MPRA

Munich Personal RePEc Archive

Voluntary Contributions by Consent or Dissent

Tan, Jonathan H.W.; Breitmoser, Yves and Bolle, Friedel

University of Nottingham, UK, EUV Frankfurt (Oder)

09. April 2010

Voluntary Contributions by Consent or Dissent[*]

Jonathan H.W. Tan[†]
University of Nottingham

Yves Breitmoser
EUV Frankfurt (Oder)

Friedel Bolle
EUV Frankfurt (Oder)

April 9, 2010

Abstract

We study games where voluntary contributions can be adjusted until a steady state is found. In *consent games* contributions start at zero and can be increased by consent, and in *dissent games* contributions start high and can be decreased by dissent. Equilibrium analysis predicts free riding in consent games but, in contrast, as much as socially efficient outcomes in dissent games. In our experiment, inexperienced subjects contribute high in consent games and low in dissent games, but behavior converges toward equilibrium predictions over time and eventually *experienced* subjects contribute as predicted: low in consent games and high in dissent games. Observed deviations from equilibrium in consent games are best explained by level-k reasoning, and those in dissent games are best explained by hierarchical reasoning formalized as *nested logit equilibrium*.

JEL Classification Codes: C44, C71, H41

Keywords: public good, contribution game, bounded rationality, mechanism

[†]Please direct correspondence to this author. Nottingham University Business School, University of Nottingham, Nottingham, UK, jonathan.tan@nottingham.ac.uk

1 Introduction

There is longstanding interest in how the structure of a game affects the level of co-operation, in particular voluntary contributions to public goods. An area of research concerns dynamic games where the "stock" of cooperation is gradually accumulated over a span of time, and actions are irreversible. For example, donations to fundraising drives build up over weeks, months or years, and donations cannot be withdrawn. Countries that adopt the consent system of organ donation legislation, e.g. Germany, have donor lists populated by people who opt in, and opting out thereafter requires administrative costs. Admati and Perry (1991), Compte and Jehiel (2003), Fershtman and Nitzan (1991), and Marx and Matthews (2000) studied the accumulation of contributions to public goods and projects, while Gale (2001), Lockwood and Thomas (2002), and Compte and Jehiel (2004) studied general accumulation games with positive spillovers. These studies typically find that, in the context of public goods, the accumulation process involves delay and outcomes are socially inefficient.

On the flip side, there are situations where cooperation can be decumulated. A major concern in fundraising is that donation pledges to projects—initiated at a cost on this pretext—are reneged, as discussed by Budig et al. (1992) and Loftin (1999). An alternative to the consent system of organ donation is that of dissent, e.g. used in Austria, where one is a candidate donor by default unless he opts out. Legislators must compare the efficacies of these rival systems prior to adoption or transition across systems (Matesanz, 1998), e.g. the UK Department of Health recently considered this issue (OTF, 2008). Further examples of analogous problems include the conservation or exploitation of common pool resources (Bolle, 1986), and the use or abuse of group reputation, e.g., eco-labeling (Hamilton and Zilberman, 2006).[1] The decumulation problem has not been systematically studied yet, and so we seek to take an initial step towards filling this gap.

---

[1] Operating in accordance with the jointly optimal exploitation path of a common oil field by symmetric oil producers is an initial state from which attempts to get a larger share of oil is similar to decreases in voluntary contributions. Members of renowned broker associations, restaurants with acknowledged quality, or holders of eco-labels for green food or environmentally friendly production can comply to the standard. Alas, irreversible decreases in group standards, and in turn common reputation, can result from opportunism (e.g. by decreasing costs through bad practice).

This paper presents a theoretical and experimental investigation of the decumulation problem ("dissent game") in contrast to the accumulation problem ("consent game"). We focus on the context of public good contributions, where each contribution yields positive spillovers to all players, but the return to each player is less than the private cost incurred. A game starts with each player having an *endowment* for financing contributions, and with *default pledges* exogenously set to zero in consent games and equal the endowment in dissent games. In each round, players simultaneously decide on whether to maintain their *interim pledges*, or to increase (decrease) them in consent (dissent) games. Interim pledges are perfectly observable. A game ends when players mutually acquiesce, i.e. pledges are no longer adjusted, or when pledges can no longer be adjusted, i.e. when pledges hit the ceiling (floor) in consent (dissent) games. Payoffs accrue from the returns from the *final pledges* standing at the end of the game, plus the remaining endowment.

The "opt-in" consent game is a monotone game with positive spillovers (Gale, 2001). Here, players have an incentive to increase their contribution only if doing so triggers off contributions by others. With budget constraints and discrete action spaces, sequences of contribution increments are finite. In turn, no such sequence is self-sustaining: the last mover in any sequence is best off deviating since his increment does not trigger off further increments.[2] The "opt-out" dissent game has not been analyzed until now.[3] We show that for specific rates of return to public good contributions, all but one of the equilibria are implausible in that they are both risk dominated and Pareto dominated by a specific alternative equilibrium. This alternative is uniquely Pareto perfect, i.e. renegotiation proof in the sense of Farrell and Maskin (1989), and also the unique limit point of logit equilibria.[4] Thus, for a range of rates of return to contributions, efficient contributions result without delay. For rates beyond this range, contributions are inefficient but still positive.

---

[2]In the threshold public goods game of Marx and Matthews (2000), only aggregate contributions are observable, and so the last player in the move sequence contributes to complete the project; inefficiency results if completion does not require the full commitment of all players. Another variant with different results is by Lockwood and Thomas (2002), who assumed continuous action spaces.

[3]The closest relative is a game by Romano and Yildirim (2005) with a fixed and finite time horizon, which is outcome equivalent to two-round games and zero contributions result in all equilibria.

[4]This theoretical result for our dissent game is interesting in that Choi et al. (2008) have found logit equilibrium to be a good explanation for their experimental data on monotone (consent) games.

In theory, consent and dissent games yield very different outcomes. We test the extent to which this bears out in the experimental laboratory, and then explain how subjects deviate from equilibrium. Doing so can further our understanding of behavior in real world scenarios with similar structures, and can guide the design of mechanisms to be used in organizations, fundraisers, or public sector solutions. Our experimental analysis involves three-player games where the quality of the public good, i.e. the rate of return to contributions, α, is either low or high. In consent games, zero contribution results regardless of α. In dissent games, social efficiency is predicted for only one of the two αs considered. We further test treatments that weaken the irreversibility constraint by allowing players to restart the dissent game at a cost, when all players cast their votes to do so. Induction yields largely similar equilibria as those for the games with perfect irreversibility. To test learning effects, we compare two information conditions, one where subjects get feedback on co-players' votes on restarting the game, and the other where feedback is not provided.

We find that in the first half of the experiment, mean contributions start off high in consent games and low in dissent games, but converge toward their respective equilibrium predictions: contributions fall in consent games and rise in dissent games. Behavior during the second half of the experiment is relatively stable, with cooperation rates staying at low levels for consent games and at high levels for dissent games. Weakening irreversibility in dissent games promotes more efficient contributions in the treatment with feedback on restart votes. Contributions in dissent games should, in theory, be high only for one of the α values, but we find that they are high regardless of α. To understand why, we estimate models of strategic choice that specify how subjects deviate from the predictions.

Logit equilibrium, originally defined by McKelvey and Palfrey (1995, 1998), cannot explain this behavior in dissent games—unlike their fit in threshold monotone games as observed by Choi et al. (2008). We examine alternative explanations such as lack of equilibration (level-k reasoning) and violations of IIA (hierarchical reasoning) formalized as *nested logit equilibrium* (NLE).[5] To the best of our knowledge, examining these concepts in the context of dynamic games is novel. Level-k

---

[5]Nested logit equilibrium is based on the "nested logit model" of choice (McFadden, 1978), and is defined as a strategy profile where all players play mutual nested logit responses.

reasoning explains behavior in consent games best. In contrast, we find that subjects' strategies in dissent games are best described by NLE, i.e. subjects have a tendency to match their opponents' pledges as opposed to playing purely payoff-monotonic logit responses. Nested logit strategies explain particularly clearly why contributions are largely independent of α. Bounded rationality, in the sense of level-*k* and NLE for the consent and dissent games, respectively, diminishes over time and this explains the observed convergence toward equilibrium.

Section 2 contains the theoretical analysis. Section 3 describes the experiment. Section 4 reports the experimental results. Section 5 discusses the comparative ability of strategic choice models to explain the data. Section 6 concludes.

2 The theory

2.1 Notation

The set of players is denoted as $N = \{1, \ldots, n\}$, typical players are $i, j \in N$. The players make contributions $a_i \in A_i$ to a public good, and the set of possible contribution levels $A_i \subset \mathbb{R}$ is finite.[6] Profiles of contributions are denoted as $a \in A := \times_{i \in N} A_i$, and i's payoff function is $p_i : A \to \mathbb{R}$. The payoffs are decreasing in the own contribution and increasing in the opponents' contributions. Thus, for all $i \neq j$,

$$a_i > a_i' \qquad \Rightarrow \qquad p_i(a_i, a_{-i}) < p_i(a_i', a_{-i}), \qquad (1)$$

$$a_j > a_j' \qquad \Rightarrow \qquad p_i(a_j, a_{-j}) > p_i(a_j', a_{-j}). \qquad (2)$$

These restrictions imply that the minimal contribution $\min A_i$ is uniquely optimal in response to all a_{-i}, i.e. rational players make minimal contributions. A well-known example satisfying Eqs. (1) and (2) is the linear public-goods game, i.e.

$$p_i(a) = \alpha * \sum_{j \in N} a_j - a_i \qquad \text{with} \quad \alpha \in (0, 1). \qquad (3)$$

---

[6]The finiteness assumption is satisfied if smallest contribution units and budget constraints exist. Notably, by avoiding the continuity assumption, our framework also includes cases where players contribute by doing specific (discrete) favors or by giving specific (discrete) objects.

In linear public-goods games, all players incur the same individual costs from contributing to the public good, they benefit equally from it, and the marginal return to each contribution is constant. Neither of these assumptions is made in the following analysis, however. Note also that we treat cases where contributing is socially efficient and cases where it is socially inefficient ($\alpha < 1/n$ and $\alpha > 1/n$, respectively) in a unified manner.

We consider two types of games where "pledges" of contributions can be adjusted for an *endogenous* number of rounds. In the *consent game*, the initial pledges are $(\min A_i)_{i \in N}$ and the pledges may be increased, but not be decreased in the course of the game. In the *dissent game*, the initial pledges are $(\max A_i)_{i \in N}$ and the pledges may be decreased, but not be increased in the course of the game. Players move simultaneously in each round, and the game ends when all pledges remain unchanged for a round. To clarify the formal differences between the two extensive forms, let us define their respective sets of possible histories explicitly, denoted as H^+ in consent games and as H^- in dissent games.[7]

$$H^+ = \left\{ \left(a^1, \ldots, a^k\right) \in A^k \mid k \in \mathbb{N}_0 \text{ and } \min A < a^1 < \cdots < a^{k-1} \leq a^k \right\}, \quad (4)$$

$$H^- = \left\{ \left(a^1, \ldots, a^k\right) \in A^k \mid k \in \mathbb{N}_0 \text{ and } \max A > a^1 > \cdots > a^{k-1} \geq a^k \right\}. \quad (5)$$

Now fix $H \in \{H^+, H^-\}$. The action set of $i \in N$ after history $h \in H$ is

$$A_i(h|H) = \left\{ a_i \in A_i \mid \left(h, (a_i, a_{-i})\right) \in H \text{ for some } a_{-i} \in A_{-i} \right\}. \quad (6)$$

The probability that i chooses $a_i \in A_i(h|H)$ is denoted as $\sigma_i(h|H)(a_i)$. Let $\pi_i(\sigma, h|H)$ denote i's expected payoff under σ in subgame h of game H.

Our analysis determines the limit points of *logit equilibria* (McKelvey and Palfrey, 1995, 1998) as the precision λ approaches infinity. We switch to considering finite precision when the experimental outcomes are analyzed. To define logit equilibria, we require the notation $\bar{\pi}_i(a_i, \sigma|h)$ for the expected payoff when i plays the pure action a_i after history h and everything else develops according to σ.

Definition 2.1 (Logit equilibrium). A strategy profile σ is a logit equilibrium if all

---

[7]We write $a \geq a'$ iff $a_i \geq a'_i \ \forall i \in N$, and $a > a'$ iff $a \geq a'$ and $a' \not\geq a$.

players play mutual logit responses in all subgames, i.e. if for some $\lambda \in \mathbb{R}_+$,

$$\sigma_i(h)(a_i) = e^{\lambda \bar{\pi}_i(a_i, \sigma | h)} / \sum_{a_i' \in A_i(h|H)} e^{\lambda \bar{\pi}_i(a_i', \sigma | h)} \qquad \forall i \in N \; \forall h \in H \; \forall a_i \in A_i(h|H). \quad (7)$$

A strategy profile σ is a limiting logit equilibrium (LLE) if it is the limit of logit equilibria as λ tends to infinity. All such limit points of logit equilibria are sequential equilibria (McKelvey and Palfrey, 1998).

2.2 Analysis

We begin with the consent game ($H = H^+$). There is a unique SPE (hence a unique limiting logit equilibrium), and along the equilibrium path, the players stick to the minimal contributions. The intuition is rather simple. In a consent game, increasing one's contribution above the minimal level is rationalizable only if one's increment triggers increments of other players' contributions. Hence, a sequence of increments may be self-sustaining under subgame perfection only if it is perpetual. This is possible only if either budgets are unlimited or increments may be infinitesimal, and in any case only if players may keep increasing their contributions for an infinite number rounds (the steady state of contributions is reached only approximately). In turn, it is not possible in the kind of contribution games that we consider in this study (for a more comprehensive discussion, see Lockwood and Thomas, 2002). Here, perpetual sequences are impossible and hence contributions remain at their initial (minimal) levels in equilibrium.

Proposition 1 (Consent games). *In consent games, the SPE is unique. According to it, all players stick to their standing pledges of contributions in all subgames.*

All proofs are relegated to the appendix.

We now analyze the dissent game ($H = H^-$). Under subgame perfection, all individually rational outcomes can be sustained in SPEs. To see this, let A^* denote the set of individually rational outcomes,

$$A^* = \{a \in A \mid \forall i \in N : p_i(a) \geq p_i(\min A)\}, \quad (8)$$

and consider the following strategy profile, which sustains any $a^* \in A^*$ along the path of play.

1. all players $i \in N$ play a_i^* in the first round and stick to it in the second round

2. after any deviation from part 1, all players $i \in N$ play $\min A_i$

Deviations from part 1 trigger part 2, and deviations from part 2 are not profitable since all players move simultaneously. Hence, this strategy profile is an SPE. In general, these SPEs are not admissible, however. For, it is weakly dominated to immediately switch to minimal contributions $\min A_i$ after the first deviation from part 1. As an example, consider public goods games with integer contribution levels $A_i = \{0, 1, \ldots, 10\}$ for all $i \in N$, and assume the players anticipate to coordinate on $(10, \ldots, 10)$. Here, it is weakly dominated to immediately move down to $(0, \ldots, 0)$ if one player deviated from $(10, \ldots, 10)$. For each player, it is weakly dominant to try to coordinate on $(9, \ldots, 9)$ with the others, or on $(8, \ldots, 8)$, $(7, \ldots, 7)$, and so on.

In this spirit, one may argue that "reasonable" players coordinate on the next-best *steady state* when the previous one has been passed.[8] Such players would invariantly make the best they can, given the circumstances and the constraints of strategic equilibrium, i.e. they coordinate on Pareto efficient Nash equilibria in all subgames. Subgame perfect equilibria satisfying this property are called Pareto perfect (Farrell and Maskin, 1989), and they seem to be rather plausible solution concepts for dissent games. If restricted to pure strategy equilibria, the set of steady states sustained under Pareto perfection can be backward induced as follows. First, $a^0 = \min A$ is a steady state by definition (of dissent games). Now, if A' denotes the set of all states that Pareto dominate a^0, all states in $\min A'$ must be steady states,[9] for sticking to any $a' \in \min A'$ is the unique Pareto efficient continuation equilibrium once a' has been reached (any deviation triggers off a^0, which is Pareto dominated). Define $A^1 := \min A'$. Next, let A'' denote all states that Pareto dominate all $a' \in A' \cup \{a^0\}$. Again, all states in $\min A''$ must be steady states, and so on. This backward induction implies that the set of steady states can be defined implicitly as the (unique) set $A^s \subseteq A$

---

[8]Given a history $h \in H$, the *state* $\omega(h)$ is the profile of contributions pledged in the previous round (or $\max A$ in the first round). Given a strategy profile σ, a state $\omega \in \Omega$ is a *steady state* if for all $h \in H$ leading to ω, all players stick with probability 1 to the state contributions ω_i.

[9]Using $\min A' = \{a' \in A' \mid \nexists a'' \in A' : a'' < a'\}$. Note that $\min A'$ need not be unique, since $<$ is a partial ordering on A.

such that for all $a \in A$,

$$a \in A^s \qquad \Leftrightarrow \qquad a \text{ Pareto dominates all } a' \in A^s : a' < a. \qquad (9)$$

Let us now return to the intuition that "reasonable" players coordinate on the next-best steady state in all subgames. If a denotes the players' current pledges of contribution (i.e. the state), then they would coordinate on one of the Pareto efficient steady states left in reach, i.e. on the Pareto frontier of $\{a' \in A^s | a \leq a'\}$. In general, the next-best steady state is not unique. Our next result establishes that it is unique in all two-player games, and that if at least $\max A^s$ is unique, then $\max A^s$ is Pareto efficient in A. The former may be of independent interest, in light of Prop. 2, and as discussed below also interesting in relations to the three-player games we study.

Lemma 1. *Define the set $A^s \subseteq A$ such that condition* (9) *holds true for all $a \in A$.*

1. *If $|N| = 2$, then the uniqueness property applies, i.e.*

$$\forall a \in A : \qquad \max \{a' \leq a \mid a' \in A^s\} \quad \text{is a singleton.} \qquad (10)$$

2. *If $\max A^s$ is a singleton, then $a^* \in \max A^s$ is Pareto efficient in A.*

Our next main result shows that if the uniqueness property applies, then all limiting logit equilibria imply that the players act Pareto perfectly: they coordinate on the next-best steady state in all subgames. Point 2 of Lemma 1 implies that the outcome must then be on the Pareto frontier of the whole possibility set. Point 1 of Lemma 1 implies that this condition is satisfied in all two-player games.

Proposition 2 (Dissent games)**.** *If the uniqueness property applies, i.e. Eq.* (10)*, then A^s is the set of steady states in all limiting logit equilibria, and the outcome along the equilibrium path is $a^* \in \max A^s$.*

Thus, there are two crucial ingredients required for efficiency in dissent games: the uniqueness property and limiting logit equilibrium. If the uniqueness property applies, then limiting logit equilibrium refines subgame perfection to the effect that the outcome is unique, renegotiation proof (in the sense of Pareto perfection), and Pareto efficient. In such cases, dissent games induce socially efficient contributions.

Table 1: A three-level public goods game

	y_1	y_2	y_3
x_1	$5,5$	$2,6$	$-1,7$
x_2	$6,2$	$3,3$	$0,4$
x_3	$7,-1$	$4,0$	$1,1$

The requirements are rather demanding in the formal sense for a game with such a simple structure—players do not move down because opponents would follow suit.

On the one hand, trembling-hand perfection does not suffice, i.e. we require equilibrium refinement beyond perfection. As an illustration, Table 1 depicts the payoffs of a game with two players and three levels of contributions each. Player 1 chooses from (x_1, x_2, x_3), of which x_1 induces the highest contribution, and player 2 chooses from (y_1, y_2, y_3), of which y_1 induces the highest contribution. The last possible state that Pareto dominates the minimal contributions (x_3, y_3) is (x_2, y_2), and the only state that Pareto dominates (x_2, y_2) is (x_1, y_1). These three states are the only steady states in limiting logit equilibria, and (x_1, y_1) results along the equilibrium path. However, both players moving toward (x_2, y_2) in the first round and sticking to it is part of a perfect equilibrium.[10] The difference to logit equilibrium is that implausible trembles may be infinitely more likely than plausible ones—where plausibility is taken in terms of expected payoffs—while limiting logit equilibria (and proper equilibria, for that matter) restrict the relative tremble probabilities. Implicitly, our experiment investigates whether subjects conform to these restrictions.

On the other hand, the uniqueness property cannot be considered a mild assumption. To illustrate this, Table 2 depicts a linear public goods game, Eq. (3), with three players and two levels of contributions per player (high: $a_1 = a_2 = a_3 = 1$, and low: $a_1 = a_2 = a_3 = 0$). If $\alpha < 1/3$, then the public good is socially not desirable. If $\alpha \in (1/3, 1/2)$, then the uniqueness property applies, since $(1, 1, 1)$ is the unique state

---

[10]To see that it this is perfect, assume player 2 plays y_3 with probability ε and y_1 with ε^2 whenever either of these options is available. If player 1 deviates to x_1 in the first round, then the probability that the state (x_1, y_1) results increases by ε^2, and the probability that the state (x_3, y_3) results increases by ε—the net effect is negative as ε tends to zero.

Table 2: A three-player, two-level public-goods game

$$a_3 = 1 \qquad\qquad\qquad\qquad a_3 = 0$$

	$a_2 = 1$	$a_2 = 1$			$a_2 = 1$	$a_2 = 0$
$a_1 = 1$	$\begin{pmatrix} 3\alpha - 1 \\ 3\alpha - 1 \\ 3\alpha - 1 \end{pmatrix}$	$\begin{pmatrix} 2\alpha - 1 \\ 2\alpha \\ 2\alpha - 1 \end{pmatrix}$		$a_1 = 1$	$\begin{pmatrix} 2\alpha - 1 \\ 2\alpha - 1 \\ 2\alpha \end{pmatrix}$	$\begin{pmatrix} \alpha - 1 \\ \alpha \\ \alpha \end{pmatrix}$
$a_1 = 0$	$\begin{pmatrix} 2\alpha \\ 2\alpha - 1 \\ 2\alpha - 1 \end{pmatrix}$	$\begin{pmatrix} \alpha \\ \alpha \\ \alpha - 1 \end{pmatrix}$		$a_1 = 0$	$\begin{pmatrix} \alpha \\ \alpha - 1 \\ \alpha \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

that Pareto dominates zero contributions. The uniqueness property does not apply if the public good's quality α is good, however. For $\alpha \in (1/2, 1)$, the minimal states that Pareto dominate $(0, 0, 0)$ are $(1, 1, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$. In turn, $(1, 1, 1)$ does not Pareto dominate these states, i.e. $(1, 1, 1)$ is not a steady state.

In this case, a coordination issue arises: Which state (i.e. which equilibrium) would the players coordinate on? Besides the three pure equilibria, there are several mixed ones. In the symmetric equilibrium, all players randomize and move down to zero contributions with probability $\sigma = \frac{4\alpha - 2}{3\alpha - 1}$ in the first round. If $\alpha = 0.6$, for example, each player moves down with probability $1/2$, and the expected payoffs are $(1/3, 1/3, 1/3)$. This symmetric equilibrium is not Pareto dominated by the other equilibria, and hence it is Pareto perfect although it is not Pareto efficient overall. This exemplifies that Pareto perfection does not generally imply Pareto efficiency, and thereby underlines the strength of Prop. 2, and even more strikingly, it emphasizes the technical relevance of the uniqueness property. For, if the uniqueness property does not apply, then Pareto perfection is not socially efficient in dissent games, and hence dissent games may exhibit significant inefficiency (although expected contributions remain positive, contrary to consent games). Surprisingly, this applies if α is too high. In this sense, a public good can be "too good".

Our experiment tests the empirical relevance of the uniqueness property. Furthermore, the zero contribution state $(0, 0, 0)$ results with probability $3\sigma^2(1 - \sigma) + \sigma^3 = 0.5$ in the mixed equilibrium, and in this case the donors end up in a state where they would renegotiate ex post if possible. Our experiment will also allow for such renegotiation by conducting treatments where "restarts" are possible.

11

3 Experimental design

3.1 The treatments

In all treatments, games of three players are considered. The possible levels of contributions (in "Euros") to the public good are $A_i = \{0, 1, \ldots, 10\}$ for all $i \in N$, the payoff function is linear, see Eq. (3), and the return factors are either $\alpha = 0.4$ or $\alpha = 0.6$. Both upward (H^+) and downward (H^-) games are considered. Combined, this 2×2 set of treatments will be referred to as the four *basic treatments* in the following. In the two additional *renegotiation treatments*, the players played dissent games with either $\alpha = 0.4$ or $\alpha = 0.6$, and when all pledges remained unchanged for a round, they were asked if the game should be restarted (for a fee of 1 Euro each, deducted from the final outcome). The game is restarted if and only if all players vote in favor of doing so. In one treatment, we provided subjects with feedback on the votes of others in the group, but in the other, there was no such feedback. We expected the treatment with feedback on votes to enhance learning. Such feedback could be informative and instructive as it allows demonstrations of the desire to cooperate, e.g. when a subject votes for a restart after decreasing his contribution to zero (to remove the disincentive for free riders to not vote for a restart if, instead, he maintained a positive contribution)—even if the restart was unsuccessful.

The equilibrium predictions are as follows. In the consent games, the players make zero contributions along the path of play (Prop. 1). In the dissent games with $\alpha = 0.4$, the uniqueness property applies, and thus players make Pareto efficient contributions in equilibrium (Prop. 2). The backward induction generating the set of steady states, see Eq. (9), implies that $a \in A$ is a steady state if and only if a is symmetric.

$$A^s(\alpha = 0.4) = \{a \in A \mid a_1 = a_2 = a_3\} \tag{11}$$

If $\alpha = 0.6$, it is sufficient that two players contribute symmetrically, i.e. all states where one player contributes nothing and the other two make symmetric contributions are steady states. By definition of A^s, the only other steady state is $(a_i, a_j, a_k) = (1, 5, 6)$ for $i \neq j \neq k$. That is, the set of steady states is

$$A^s = \{a \in A \mid \exists i, j, k : a_i = 0 \wedge a_j = a_k \text{ or } (a_i, a_j, a_k) = (1, 5, 6)\} \tag{12}$$

This set of steady states is sustained also in the unique limiting logit equilibrium. The initial state $(10, 10, 10)$ is therefore not stable, and in the symmetric equilibrium, all players randomize over reducing their contributions. The experiment will inform us to which degree the formal relevance of the uniqueness property translates into practical relevance. The possibility of ex-post renegotiation is theoretically relevant only in case $\alpha = 0.6$, where the event that all players reduce their contributions has positive probability in limiting logit equilibria.

3.2 Experimental procedure

The experiment was conducted in a purpose-built experimental economics laboratory in Germany, where total of 204 subjects participated. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). There were two main treatments, consent (U) and dissent (D), as described in section 2. We label the treatment with the restart option without feedback about co-participant's votes (DR) and the one with feedback (DRF). There were twelve subjects in each of the 17 sessions, who played U, D, DR, or DRF. Each session was (virtually) partitioned into two independent sub-sessions of six subjects each. Thus, we had a total of eight independent observations for U, ten for D, eight for DR, and eight for DRF. This design enhances the perceptions of anonymity between subjects, given large session sizes, while yielding more independent observations.

Each session comprised 12 stages (games). In each game, subjects were randomly re-matched into groups of three to implement the one-shot game. This allows for experience, and eliminates reputation or supergame effects. In each session (treatment), the quality of the public good, i.e. the return rate to contributions α, was either 0.4 or 0.6 in either the first or second half of the experiment (i.e. games 1-6 or 7-12, respectively). This within-subject design allows us to test the robustness of behavior in each treatment when subjects are exposed to different return rates.

At the start of each game, each subject received an endowment of 10 Euros, and the individual pledge of each subject was set at the default of 0 (10) Euros in the U (D, DR, or DRF) treatment. In each round, subjects could increase (decrease) their contributions in the U (D, DR, or DRF) treatment, if they were not already

contributing everything (nothing). After each round, subjects received feedback about the current contributions of each co-player in the group, which is especially important for dissent games with $\alpha = 0.6$, where there are multiple equilibria and coordination is a concern. The game continued as long as at least one subject had adjusted the pledge in the stage, and stopped when all players in the group simultaneously acquiesced for one stage. At the end of the game, all subjects were informed of their earnings.

At the beginning of the experiment, subjects were randomly seated. They were then required to read the experimental instructions, provided on printed sheets, and to answer a short control questionnaire for us to check their understanding. We used neutral language throughout. Subjects in doubt were verbally advised by the experimental assistants before being allowed to begin. Computer terminals were partitioned, so that subjects were unable to communicate, gesture, or to look at other computer screens. Decisions were thus made in privacy. At the end of the experiment, subjects were informed of their payments, and asked to privately choose a code name and password. This was used to anonymously collect their payments from a third party one week after the experiment. Each subject received participation fee of 6 Euros and the earnings from one randomly chosen "winning stage". The average payment per subject was 19 Euros for approximately 1.5 hours per session.[11]

4 Experimental results

4.1 Aggregate behavior across treatments and time

This first part of our analysis compares, across treatments, the cooperation levels and trends over time. We analyze behavior overall, and when subjects are inexperienced (block 1, games 1–6) or experienced (block 2, games 7–12). Figure 1 presents the mean contributions over time for each treatment. Table 3a summarizes the mean contributions over all games, the first game and the final game, in each block, followed by Table 3b showing the univariate test results of pairwise comparisons of mean con-

---

[11]The monetary incentives provided are substantial by local standards. The average payment is, for example, 50% more than the wage of a research assistant in this university.

tributions between treatments for all rounds in each block.[12]

Table 3c presents the results of ordered logit regressions that test the marginal effects of treatment differences on contributions, while controlling for experience and α. The treatment variables are α (= 0.4 or = 0.6), D (= 1 for all dissent games, or = 0 if otherwise), and R (= 1 for DR, and = 0 if otherwise), RF (= 1 for DRF or = 0 if otherwise). We also include the time variable G (= 1 to 12 corresponding to the respective game number), and interactions of treatment variables D, R, and RF by G. We have chosen discrete (ordinal) regression, because there are only 11 discrete contribution levels per subject, and proportional odds modeling (McCullagh, 1980) in particular, because the specific contribution levels have cardinal interpretability.

Inexperienced play   In block 1, the univariate tests reported in Table 3b show that, for both αs, overall mean contributions are not significantly different across treatments. At the outset of the experiment, i.e. in game 1, mean contributions are significantly higher in consent games than in dissent games (see Table 3c: the coefficient of D is significantly negative in Block 1), and the DRF treatment is initially even worse than in the D treatment. This difference vanishes by game 6 when $\alpha = 0.4$ (all differences are insignificant), and contributions in game 6 are significantly higher in the downward treatments when $\alpha = 0.6$ (except DR where the difference is insignificant). Univariate tests comparing round-specific contributions concur. Similar to the decay observed in most public good experiments, contributions in the consent treatments decrease over time (see that G is significantly negative overall). In stark contrast, those of the dissent treatments increase over time ($D \times G$ is significantly positive in Block 1), and the increase is accelerated in the DRF treatments ($RF \times G$ is significantly positive, too).

Experienced play   In block 2, mean contributions are stable over time—the time interaction variables are insignificant in all cases. According to our Bayes Information Criterion (BIC) measures, these interactions should be excluded in the regression

---

[12]We used Mann-Whitney U tests, and define an independent observation as the average of each session. Inter-treatment comparisons of mean contributions in games 1, 6, 7 and 12 are reported in Table 10 of the supplementary material.

Table 3: Mean contributions over time

(a) Mean contributions overall and in first and final rounds of each block

Inexperienced	α = 0.4			α = 0.6		
	Overall	Game 1	Game 6	Overall	Game 1	Game 6
U	3.36	3.88	2.75	4.01	4.92	3.38
D	1.86	0.93	2.60	5.93	3.13	7.43
DR	1.83	0.58	2.83	4.36	2.04	5.75
DRF	3.24	1.17	4.08	4.48	1.33	7.79

Experienced	α = 0.4			α = 0.6		
	Overall	Game 7	Game 12	Overall	Game 7	Game 12
U	1.91	2.63	1.54	4.65	5.13	3.79
D	6.31	6.77	6.77	6.38	5.97	6.47
DR	4.50	4.83	5.08	5.63	5.67	6.00
DRF	6.44	6.58	5.50	7.21	4.92	7.29

(b) Nonparametric tests of mean contributions and *p*-values (in parentheses) for α = 0.4 (in cells below diagonal) and α = 0.6 (in cells above diagonal)

Block 1	*U*	*D*	*DR*	*DRF*
U	–	−1.599 (0.1099)	0 (1)	−1.155 (0.2482)
D	1.470 (0.1416)	–	−0.735 (0.4624)	−0.735 (0.4624)
DR	1.732 (0.0833)	0 (1)	–	−0.289 (0.7728)
DRF	0 (1)	−1.470 (0.1416)	−1.443 (0.1489)	–

Block 2	*U*	*D*	*DR*	*DRF*
U	–	−0.735 (0.4624)	−1.155 (0.2482)	−1.155 (0.2482)
D	−2.449 (0.0143)	–	0.245 (0.8065)	0.245 (0.8065)
DR	−1.597 (0.1102)	0.980 (0.3272)	–	−1.443 (0.1489)
DRF	−2.309 (0.0209)	−0.490 (0.6242)	−1.016 (0.3094)	–

(c) Ordered logit regression (proportional odds) of final contributions

α	*D*	*R*	*RF*	*G*	*D* × *G*	*R* × *G*	*RF* × *G*
Blocks 1 + 2							
4.157** (0.368)	−0.952** (0.201)	−0.229 (0.216)	−0.356 (0.218)	−0.046* (0.019)	0.246** (0.028)	−0.050 (0.030)	0.082** (0.031)
Block 1							
6.058** (0.532)	−1.508** (0.312)	−0.141 (0.333)	−0.842* (0.344)	−0.099 (0.057)	0.411** (0.082)	−0.105 (0.086)	0.234** (0.088)
Block 2							
2.457** (0.516)	1.231** (0.140)	−0.633** (0.144)	0.275 (0.149)	0.044 (0.030)			

Note: Two-tail *p*-values at the 5% (1%) level are denoted by * (**). The 10 intercepts per regression model are reported in the supplementary material (Table 8).

Figure 1: Mean contributions over time
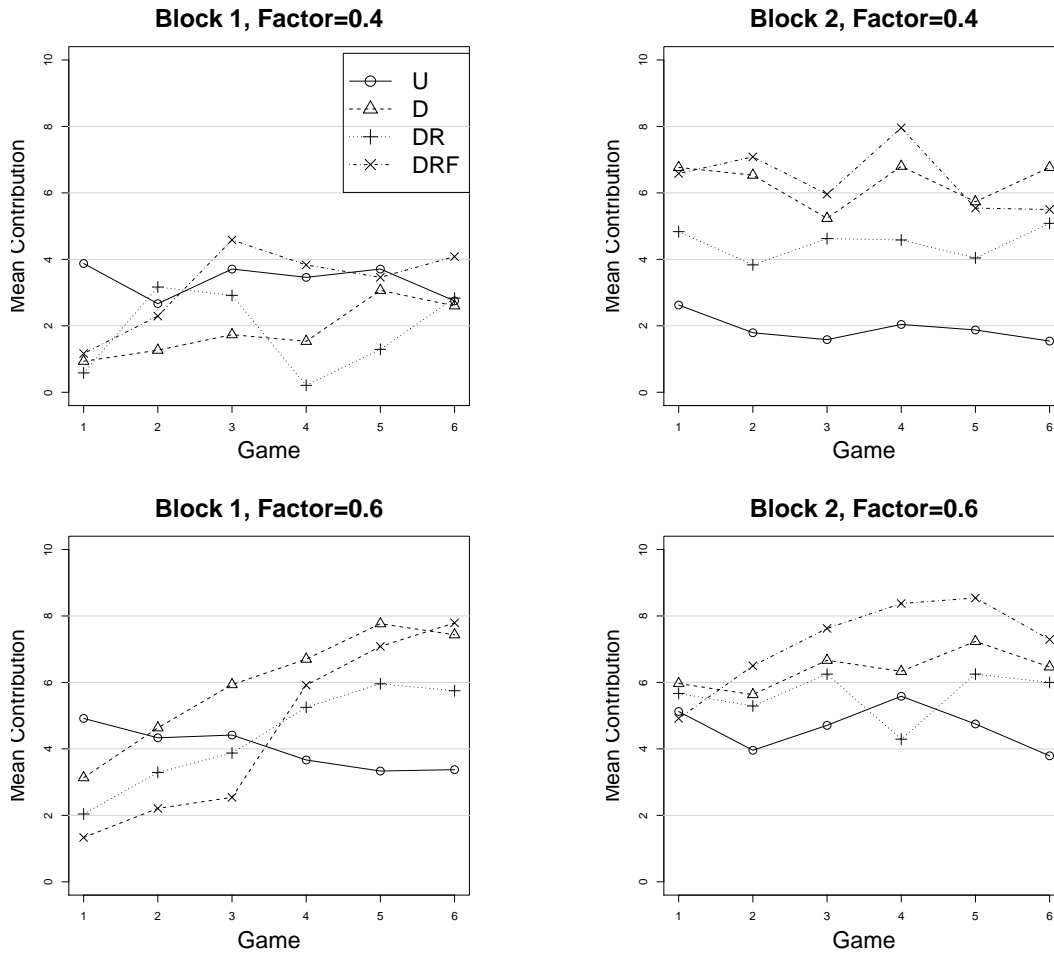
model for block 2 (see also Table 8 in the supplementary material). Table 3c shows that α has less impact on experienced than inexperienced behavior (α is 2.457 in block 2 compared to 6.058 in block 1), and that contributions are significantly higher in the dissent treatments than in *U* (see that *D* is positive and significant). There is no significant difference between *D* and *DRF* in Block 2 (i.e. *RF* is insignificant), both eliciting higher contributions than *U*. Overall contributions are significantly higher in all dissent games than in consent games when $\alpha = 0.4$ (e.g. 19% for *U* versus 63% for *D*) and they are rather similar when $\alpha = 0.6$. Once again, these results concur with the univariate inter-treatment comparisons. When $\alpha = 0.6$, the mean is 64% for *D*, and the mean contributions go as high as 80% and 85%, as observed in *DRF* games

10 and 11 when $\alpha = 0.4$ and 0.6 respectively. Even when the low return rate of 0.4 makes the public good less attractive, in some dissent treatments mean contributions are more than two or three times that found in U, e.g. 63% in D versus 19% in U.

Restart treatment Overall, restarts were rarely observed. No restarts occurred in DR, and yet, the treatment of "restarts without information" affected behavior negatively as subjects gained experience ($R = -0.633$ in block 2 is significant, see Table 3c). A mere eight games were restarted in DRF, twice in block 1 and six times in block 2 (i.e. primarily when subjects are experienced). Restarts had a dramatic effect on contributions: the mean contributions of 0 in the first try of games preceding restarts increased to 60% and 78% in the second try for $\alpha = 0.4$ and 0.6, respectively. Contributions were low at the initial game of the experiment but increased as subjects gain experience. This shows that subjects learn faster when restarts are possible, depending on whether the votes on restarting are disclosed. While contribution levels suffer in the short run from the possibility of restarts, the quick increase in contributions and eventual stability more than compensates in the long run.[13]

4.2 Behavior within games

We now scrutinize interim pledges, i.e. behavior within games. Figure 2 shows the mean deviation of pledges from the default pledge of 0 in consent treatments and from 10 in dissent treatments in round 1 of each game. Mean deviations appear to decrease over time, and more so in block 1 and in the dissent treatments. This decreasing effect more or less stabilizes in block 2, but there remains a low tendency for subjects to deviate from the default pledges.

To systematically understand the structure of interim actions, in particular how subjects respond to the moves of co-players, we estimate the subjects' contribution "strategies" as functions of the following independent variables: let $\underline{c}$ be the lesser of the opponents' standing pledges, $\overline{c}$ be the greater of their pledges, α, and D, R, RF,

---

[13]When restarts were possible, some subjects were observed to drop their contributions down to zero following "mistakes" of their co-players, who can understand the cooperative motives of this behavior only when the votes were disclosed.

Figure 2: Mean adjustments in round 1 across games
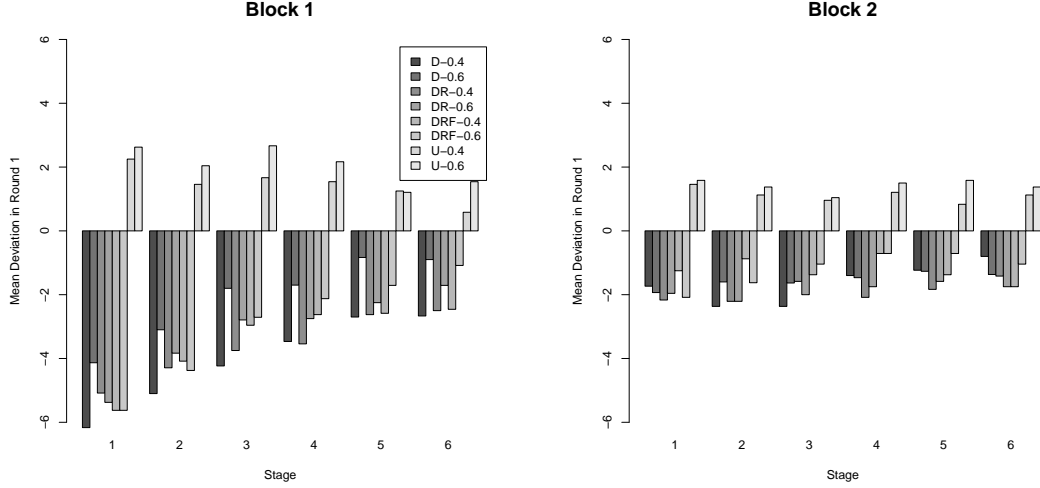
and G be as previously defined. As before, we use the proportional odds model of ordered logit regression.[14]

We discuss the strategy estimates for U and D treatments as well as block 1 and block 2 separately. The estimate for the pooled datasets, i.e. after pooling either U and D or blocks 1 and 2, are reported in the supplementary material (Table 9). Let us begin with block 1 in dissent games. The estimated strategy is[15]

$$c_i^{D,1} \mathrel{\hat{=}} I_c + \underset{(0.016)}{0.090} \cdot \underline{c}^* + \underset{(0.019)}{0.255} \cdot \overline{c}^* + \underset{(0.358)}{4.321} \cdot \alpha^* - \underset{(0.083)}{0.322} \cdot R^* - \underset{(0.086)}{0.039} \cdot RF + \underset{(0.021)}{0.242} \cdot G^*$$

with I_c as a shortcut of the 10 intercepts in proportional odds models (for $0|1$, $1|2$, ..., $9|10$ in our case), which are reported in Table 9, and $\hat{=}$ as a shortcut of the proportional-odds logit link. We will discuss this strategy in relation to the estimate for block 2, which follows.

$$c_i^{D,2} \mathrel{\hat{=}} I_c + \underset{(0.018)}{0.256} \cdot \underline{c}^* + \underset{(0.022)}{0.207} \cdot \overline{c}^* + \underset{(0.428)}{0.092} \cdot \alpha - \underset{(0.096)}{0.347} \cdot R^* - \underset{(0.111)}{0.584} \cdot RF^* + \underset{(0.024)}{0.074} \cdot G^*$$

---

[14]Due to the irreversibilities that limit behavior within games, however, we have to control for classification variability in the regression model. Initially, the subjects' choices may fall in all possible classes $0, 1, \ldots, 10$, but later in the game, if the standing pledge is c, the choices are restricted to either $0, 1, \ldots, c$ or $c, c+1, \ldots, 10$ (in dissent or consent games, respectively). This source of variability is a mere technical issue that can be resolved straightforwardly however (see also Farewell, 1982).

[15]In parentheses, we report the standard errors of the estimates. Parameters that are significant at the 1% level are marked by an asterisk (*).

19

Result 4.1 (Dissent games). *In dissent games, the greater of the opponents' pledges ($\overline{c}$) and time (G) positively affect contributions, while restart possibilities do so negatively. As subjects gain experience, the effect of public good quality (α) becomes insignificant, the effect of time (G) becomes almost insignificant, and the lesser of the opponents' pledges $\underline{c}$ gains in significance.*

The observation that *G* becomes almost insignificant suggests that behavior stabilizes in block 2, and the fact that the lesser of the opponents' pledges ($\underline{c}$) is more relevant in block 2 suggests that subjects start reasoning more strategically (e.g. in the predicted equilibrium for $\alpha = 0.4$, only $\underline{c}$ is relevant). Note also that contribution strategies are independent of α in dissent games. The negative significance of *RF* in block 2 relates to the fact that subjects learn that it would pay off to race down to zero (after "mistakes") in *DRF* treatments.

The strategy estimates for blocks 1 and 2 in consent games are

$$c_i^{U,1} \mathrel{\hat{=}} I_c + \underset{(0.042)}{0.081 \cdot \underline{c}} + \underset{(0.031)}{0.153 \cdot \overline{c}^*} + \underset{(0.545)}{0.418 \cdot \alpha} - \underset{(0.032)}{0.029 \cdot G}, \tag{13}$$

$$c_i^{U,2} \mathrel{\hat{=}} I_c + \underset{(0.049)}{0.725 \cdot \underline{c}^*} + \underset{(0.038)}{0.033 \cdot \overline{c}} + \underset{(0.662)}{3.242 \cdot \alpha^*} + \underset{(0.032)}{0.002 \cdot G}. \tag{14}$$

Result 4.2 (Consent games). *In consent games, inexperienced subjects follow the leading contributor ($\overline{c}$), while experienced subjects follow the trailing contributor ($\underline{c}$) amongst the opponents and react to the public good quality α.*

That is, experienced subjects are immune to opponents that try to "lead by example." They react to increases of the lesser of their opponents' pledges, which suggest that experienced subjects try to avoid being the bad example rather than following or being the good one. In contrast to to dissent games, α becomes significant for experienced subjects, and time *G* is insignificant.

5   Strategic choice analysis

Section 4 showed that behavior significantly deviated from equilibrium at the start of the experiment, but tends to approach the equilibrium predictions over time. This indicates that judgmental biases disappear with experience. We further investigate the

nature of bias and its effect over time by modeling strategic choice in an econometrically precise manner.

5.1 Candidate models of strategic choice

An intuitive and commonly used explanation for deviations of behavior from equilibrium is that subjects do not strictly maximize expected payoffs or utilities. The conventional choice-theoretic approach to describing this bias is random utility maximization in the form of logit response, Eq. (7), with precision $\lambda < \infty$. Our modeling approach adopts the logit equilibrium, following McKelvey and Palfrey (1995, 1998), i.e. mutual logit response. Choi et al. (2008) have shown that logit equilibrium reflects strategic choice in their monotone games fairly well. Since logit equilibria are not scale invariant, we linearly normalize the payoffs of all games so that the minimal payoff is one and the maximal payoff is ten in each case. This allows us to compare utilities across treatments. Alternatively, one can adjust the precision parameter λ between games. In addition to finite precision λ, we consider potential biases due to nonequilibrium play and violation of independence from irrelevant alternatives (IIA).[16]

Many studies have been investigating deviations from (logit) equilibrium by relaxing the assumption that subjects have equilibrium beliefs. The best known class of such models is *level-k reasoning* following Stahl and Wilson (1994, 1995), and more generally the *cognitive hierarchy* approach of Camerer et al. (2004). The idea is that level-0 players act non-strategically, e.g. by randomizing uniformly, level-1 players believe that their opponents are level 0 (and logit respond to this belief), level-2 players assume that their opponents logit respond to level 0, et cetera. In our econometric analysis, we consider the possibility that the subject pool is a mixture of up to three of such level-k types (each with precision λ_k) and up to two logit equilibrium types. For example, let $K = \{0, 1, 2, e_1, e_2\}$ denote the set of types in the population, let $\lambda = (\lambda_k)_{k \in K}$ and $\rho = (\rho_k)_{k \in K}$ denote the types' precisions and population shares, re-

---

[16]The supplementary material (Appendix C) contains an extended analysis of deviations from logit equilibrium including interdependent preferences, subjective and heterogeneous precisions λ (following Weizsäcker, 2003, and more generally Rogers et al., 2009), and limited look-ahead (or depth of reasoning) related to Kübler and Weizsäcker (2004). These biases are found to be less relevant.

spectively, and let $o_{s,t}$ denote the tth observation of subject $s \in S$. Thus, if $\sigma(o_{s,t}|k,\lambda)$ denotes the probability of $o_{s,t}$ for type $k \in K$ under the precision parameters λ, then the overall log-likelihood is

$$LL(o \mid \lambda, \rho) = \ln \prod_{s \in S} \sum_{k \in K} \rho_k \cdot \prod_t \sigma(o_{s,t}|k,\lambda). \tag{15}$$

Finally, we consider the possibility that choices are not independent from irrelevant alternatives (IIA). To see why subjects may violate IIA in contribution games, consider a dissent game with standing pledges $(8,6,6)$. Arguably, player 1 first decides whether to stay above the opponents' pledges ($c = 7,8$), to match them ($c = 6$), or to undercut their pledges ($c \leq 5$), and only secondly he chooses the pledge from the chosen category. Such choices violate IIA. For example, the probability of choosing $c = 6$ in relation to that of choosing $c = 5$ is not independent of how many choices there are in the lowest category $c \leq 5$ in this case. We utilize the concept of mutual nested logit responses, i.e. *nested logit equilibrium* (NLE), to model this two-level choice hierarchy in the context of strategic equilibrium.[17] Daly and Zachary (1978) and McFadden (1978) have shown that nested logit response is compatible with random utility maximization. The existence of NLEs in normal form games is thus implied by Theorem 1 of McKelvey and Palfrey (1995); the existence of agent NLEs in contribution games under complete information follows immediately.[18]

To formally describe how NLE extends logit equilibrium, Def. 2.1, recall the notation $A_i(h|H)$ for i's action set after $h \in H$ and $\bar{\pi}_i(a_i, \sigma|h)$ for i's expected payoff from action a_i as it was introduced prior to Eq. (7). For all $i \in N$ and $h \in H$, let $\left\{A_i^{(r)}(h|H)\right\}_{r=1}^5$ denote a partition of $A_i(h,H)$ into five possibly empty nests (e.g. as described in Footnote 17).

Definition 5.1 (Nested logit equilibrium, NLE)**.** A strategy profile σ is an NLE if all players play mutual nested logit responses in all subgames, i.e. if $\lambda \in \mathbb{R}_+$ and $\kappa \in [0,1]$

---

[17]In our model, there are five nests each depending on how a given action relates to the opponents' current pledges. The first nest contains all options below opponents' pledges, the second one is to match the lower pledge, the third one is to pick an action between their current pledges, the fourth one is to match the higher pledge, and the fifth one is to pick a pledge above both opponents' pledges.

[18]The concept of nested logit response has been employed successfully in modeling economic choice (Anderson and De Palma, 1992; Whitten and Palmer, 1996; Train et al., 1987; Lee, 1999; Shaw and Ozog, 1999). We are not aware of previous applications of NLE to dynamic games, however.

exists such that for all $i \in N$, all histories $h \in H$, all nests $r = 1, \ldots, 5$ and all $a_i \in A_i^{(r)}$,

$$\sigma_i(h)(a_i) = \frac{\exp\{\lambda \bar{\pi}_i(a_i)\}}{\sum_{a_i' \in A_i^{(r)}} \exp\{\lambda \bar{\pi}_i(a_i')\}} \cdot \frac{\exp\{\kappa J_i(r)\}}{\sum_{r'=1}^{5} \exp\{\kappa J_i(r')\}} \quad (16)$$

$$\text{where} \quad J_i(r') = \ln \sum_{a_i' \in A_i^{(r')}} \exp\{\lambda \bar{\pi}_i(a_i')\}, \quad (17)$$

using $A_i := A_i(h|H)$ as well as $\bar{\pi}_i(a_i) := \bar{\pi}_i(a_i, \sigma|h)$.

The choice probability is a product of two factors each having the multinomial logit form. The first one represents the probability of choosing a_i conditional on having chosen nest r, and the second one represents the probability of choosing nest r. The value of $J_i(r)$ is the "inclusive value" of nest r, and in our model (where utilities are rescaled by λ), the term $J_i(r)/\lambda$ is the expected utility from the choice among the actions in nest r.

Given this, we may illustrate precisely how NLEs allow for violations of IIA. Consider the symmetric state (c, c, c) in the dissent game. For simplicity assume i believes that his opponents stick with probability 1 to c (in this round) and that $(0, 0, 0)$ will result inevitably if i chooses to deviate from c in this round (the actual backward induction is more complex, of course). Thus, there are two nests, picking either $a_i = c$ or $a_i \in \{0, 1, \ldots, c-1\}$, and within each nest all actions are outcome equivalent. Denote the respective utilities as π_c and π_0. Rearranging Eq. (17), we may thus write the probability of choosing $a_i = c$ as

$$\sigma_i(c) = \frac{\exp\{\kappa \cdot \lambda \pi_c\}}{\exp\{\kappa \cdot \lambda \pi_c\} + \exp\{\kappa \ln c + \kappa \cdot \lambda \pi_0\}}. \quad (18)$$

The standard logit probability results if $\kappa = 1$, and in this case the quotient of the probability of sticking to the standing pledge (c) in relation to the probability of any other action is independent of how many actions there are in the second nest (i.e. IIA results for $\kappa = 1$). However, this implies that the absolute probability of sticking to the standing pledge c depends on its value. If c is small, then the probability of sticking to c is large, and if c is large, then the probability of sticking to it is small— simply because there are more ways in which i can deviate. The assumption that IIA is satisfied in this case seems implausible. Why should the salience of sticking to c in this context depend on whether the state is $(2, 2, 2)$ or $(10, 10, 10)$? The probability

23

of sticking to c is independent of the state in the limiting case $\kappa \to 0$, assuming λ is adapted such that $\kappa \cdot \lambda = \text{const.}$ Then, the relative probabilities of the nests are independent of the number of actions per nest—i.e. there obtains IIA at the level of nest choice. Intermediate values of κ allow us balance these two choice regimes.

5.2 Modeling results

Let us first test the relevance of these potential sources of biases from logit equilibrium. The Bayes Information Criterion (BIC, Schwarz, 1978) measures for these models are reported in Table 4.[19] Logit equilibrium, by itself, does better than the naive model (zero precision, i.e. $\lambda = 0$), but it is not the best fitting model. Overall, the most substantial bias is the deviation from IIA. Significant improvements in BIC are achieved by NLE, with the overall margin being more than 1000 points on the log-likelihood scale. That is, the data are e^{1000} more likely to occur under the nested logit model than under any other model. In comparison, a mere 10 log-likelihood points per parameter are highly significant in likelihood-ratio tests, and nested logit actually is parsimoniously parameterized in relation to the other models.

Result 5.1 (Overall fit). *The primary source of deviations from logit equilibrium is that subjects' choices are not independent from irrelevant alternatives.*

A closer look at the data provided in Table 4 shows that NLE fits best for dissent games, but does worst amongst all the models for consent games, which are better explained by level-k. This implies that in dissent games, pledges are primarily chosen hierarchically, in relation to the pledges of co-players, while in consent games, expected payoffs rather than relative pledges are more relevant. The following result is based on the the goodness-of-fits reported in Table 4.

Result 5.2 (Specific fit). *In dissent games, subjects mainly exhibit violations of IIA*

---

[19]The maximum likelihood estimates including standard errors (obtained from the information matrix) are reported in the supplementary material (Table 7). The likelihoods had been maximized jointly over all parameters (using the Nelder-Mead algorithm), in order to avoid inefficiency of the estimates as discussed by Amemiya (1978) for nested logit models and by Arcidiacono and Jones (2003) for finite mixture models.

Table 4: BICs of the various models, where $BIC = -LL + (\#Pars)/2 * \ln(\#Obs)$

Model	#Pars	Blocks 1+2					Block 1					Block 2				
		All	*U*	*D*	*DR*	*DRF*	All	*U*	*D*	*DR*	*DRF*	All	*U*	*D*	*DR*	*DRF*
Logit equilibrium																
LE	1	16057	3302	4294	4522	3950	8572	1772	2332	2253	2225	7467	1510	1965	2277	1727
Logit equilibrium in finite mixture with level-*k*																
LE + Lev 0	2	15831	3252	4212	4456	3934	8477	1749	2299	2242	2208	7365	1504	1932	2223	1728
LE + Lev 0–1	4	15709	3119	4220	4469	3944	8369	**1636**	2307	2258	2220	7072	1498	1834	2134	1650
LE + Lev 0–2	6	15718	3127	4228	4477	3952	8386	1644	2314	2266	2228	6998	1418	1842	2143	1660
LE + Lev 0–3	8	15493	**3068**	4223	4420	3870	8249	1656	2299	2211	2170	6910	**1351**	1850	2142	1656
LE × 2 + Lev 0	4	15345	3195	4055	4316	3823	8300	1737	2235	2193	2180	7282	1496	1923	2212	1695
LE × 2 + Lev 0–1	6	15290	3138	4063	4326	3829	8236	1667	2243	2204	2188	6986	1416	1841	2140	1655
LE × 2 + Lev 0–2	8	15277	3113	4071	4338	3842	8234	1657	2252	2214	2198	6950	1375	1849	2149	1665
Nested logit equilibrium																
NLE + Lev 0	3	**13921**	3985	**3313**	**3636**	**3021**	**7620**	1840	**1966**	**1990**	**1863**	**6136**	1971	**1357**	**1661**	**1185**
Benchmark concepts																
Naive model ($\lambda = 0$)	0	18050	5437	4180	4469	3963	9541	2815	2275	2242	2209	8509	2622	1905	2228	1755
Prop. odds logit regr.	17	12111	3917	3130	2961	2286	6898	2032	1903	1653	1480	5159	1903	1259	1326	834

Note: The BIC of the best performing model per subset of the data (i.e. per column) is set in boldface type.

(as in nested logit reasoning). In consent games, subjects mainly exhibit level-k reasoning.

This explains why we did not observe the predicted effect of the return rate α in dissent games. If subjects reason hierarchically, by first deciding how to position themselves in relation to co-players and then what pledge to make, the tendency that subjects match each others' pledges regardless of α naturally follows. The absence of this effect in consent games explains the lack of "leadership by example" and the socially inefficient outcomes observed.

Next, look at the goodness-of-fit of the various models in the overall perspective. The two relevant benchmark concepts we consider are the naive prediction of uniform randomization in all states, and the highly parameterized proportional odds model of ordered logit response (see Section 4). The former can be considered to represent the lower bound of how a model should score in any case, and the latter represents a suitable upper bound.

Result 5.3 (Relative fit). *In terms of the BIC, the three-parameter nested logit equilibrium covers 70% of the range spanned up by the naive prediction of uniform randomization (BIC = 18050) and the 17-parameter regression model (BIC = 12111). The best alternative model covers 46% of this range.*

Finally, let us analyze the choice-theoretic implications of experience in contribution games. We do so by comparing the parameter estimates of the best-fitting models (the NLE models) between block 1 and block 2 of the experiment. Table 5a shows that in NLE models, the share of level-0 subjects decreases from 10.8% to 3.4% between block 1 and block 2, while the the precisions remain largely constant.

Result 5.4 (Experience effects). *The main implication of experience is that the estimated share of level-0 types decreases. In NLE models, which fit dissent games best, the precision does not increase. But, in level-k models, which fit consent games best, the precision increases (besides the decline of level-0 types).*

That is, the main reason underlying the efficiency increase in block 2 is that more subjects reason strategically with experience.

Table 5: Parameter estimates (standard errors) of the best-fitting models

(a) NLE models ($1 - \rho$ is the share of level-0 players)

Blocks $1+2$			Block 1			Block 2		
λ	κ	ρ	λ	κ	ρ	λ	κ	ρ
3.635	0.243	0.976	4.249	0.374	0.892	3.683	0.228	0.966
(0.002)	(0.001)	(0.009)	($1e-4$)	($7e-05$)	(0.009)	($4e-05$)	($1e-05$)	(0.015)

(b) Finite mixture with level-k types

	Logit Eq.		Level 1		Level 2		Level 3	
ρ_0	ρ_E	λ_E	ρ_1	λ_1	ρ_2	λ_2	ρ_3	λ_3
			Block 1					
.470	0.370	2.514	0.159	4.702				
	(0.060)	(0.085)	(0.031)	(0.214)				
			Block 2					
.149	0.321	4.948	0.073	5.283	0.062	8.435	0.392	2.175
	(0.022)	(1.084)	(0.022)	(0.313)	(0.024)	(0.700)	(0.038)	(0.123)

6 Conclusion

With this paper, we formally introduced the dissent mechanism for eliciting voluntary contributions. A Folk Theorem applies under subgame perfection, and given the multitude of equilibria that are robust to trembling hand perfection, sharp predictions seemed out of reach. We showed, however, that for a range of return rates the unique Pareto perfect equilibrium (Farrell and Maskin, 1989) risk dominates and Pareto dominates all other equilibria; and that this solution is the unique limit point of logit equilibria. We then tested both consent and dissent mechanisms experimentally. The efficacy of the dissent mechanism was demonstrated by the high contributions elicited, outperforming the traditional consent mechanism.

Schelling (1960) proposed that dynamic structures allow cooperation to be mutually reinforced. In his words, "if the contribution is divided into consecutive small contributions, each can try the other's good faith for a small price" (p. 45). Indeed, there is experimental evidence of high cooperation rates in dynamic games when there are project completion bonuses (Duffy et al., 2007), when leaders have the power to

exclude members from future games (Güth et al., 2007), and when the quality of the public good is known only to the leader (Potters et al., 2007). In contrast to these studies, our setup is devoid of completion benefits, exclusionary power, and quality uncertainty, and it allows more gradualism with smoother endowments and an endogenous time horizon.[20] The data indicates that at outset of the experiment, subjects raised their pledges in response to having pledged less than the leading opponent. but such behavior diminished with experience.

In contrast, contributions in dissent games increased over time and the endgame effect was not observed. Mean contributions were as high as 85%, even in the final games of the experiment. Unlike in static games where "dissent" makes no difference (Dufwenberg et al., 2006),[21] the gradualism and observability allowed in dynamic games safeguard players against being suckered, and cooperation is sustained in the absence of "negative leadership" from free riders. Subjects indeed lowered their pledges in response to others having pledged less. As decrements occurred less frequently across games, pledges were maintained at higher levels within games.

The bounded rationality of subjects in consent games, best described as level-0 types in relation to level-k types, diminished over time. Random play in dissent games, as captured in relation to NLE, also diminished over time. Consistent with the equilibrium prediction for the consent (dissent) game, cooperation decreased (increased) as more and more subjects reasoned strategically. This conclusion is drawn from an extensive evaluation of various *a priori* plausible models of strategic choice. Applying the concept of NLE to dynamic games constitutes another novel contribution of this paper. We expect that NLE will prove useful in explaining behavior in other experimental games where a confluence of hierarchical reasoning and imprecise response is to be modeled.

As has been modeled and evidential in our study, "time is of the essence" in contribution games. While prescribing the ways by which the dissent mechanism

---

[20]Duffy et al. did not find a "leadership" effect in that cooperation was insensitive to whether or not others' contributions were observable. A reviewer of their paper suggested that the coarse endowments in their game might explain this result. Our setup had smoother endowments.

[21]Dufwenberg et al. (2006) compared one-shot simultaneous move games where players were either asked to "give" or to "take" from the public good—essentially varying the valence of the task while leaving the move structure and in turn the free riding equilibrium unperturbed.

should be practically implemented in specific contexts extends beyond the scope of this paper, our experimental treatments and results give reason and affords guidance for efforts in this respect.

References

Admati, A. and Perry, M. (1991). Joint projects without commitment. *Review of Economic Studies*, 58(2):259–276.

Amemiya, T. (1978). On a two-step estimation of a multivariate logit model. *Journal of Econometrics*, 8(1):13–21.

Anderson, S. and De Palma, A. (1992). Multiproduct firms: A nested logit approach. *The Journal of Industrial Economics*, 40(3):261–276.

Arcidiacono, P. and Jones, J. (2003). Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, 71(3):933–946.

Bolle, F. (1986). On the oligopolistic extraction of non-renewable common-pool resources. *Economica*, 53(212):519–527.

Budig, M., Butler, G., and Murphy, L. (1992). Pledges to nonprofit organizations: Are they enforceable and must they be enforced? *University of San Francisco Law Review*, 22:47–147.

Camerer, C., Ho, T., and Chong, J. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–898.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):817–869.

Choi, S., Gale, D., and Kariv, S. (2008). Sequential equilibrium in monotone games: A theory-based analysis of experimental data. *Journal of Economic Theory*, 143(1):302–330.

Compte, O. and Jehiel, P. (2003). Voluntary contributions to a joint project with asymmetric agents. *Journal of Economic Theory*, 112(2):334–342.

Compte, O. and Jehiel, P. (2004). Gradualism in bargaining and contribution games. *Review of Economic Studies*, 71(4):975–1000.

Daly, A. and Zachary, S. (1978). Improved multiple choice models. In Hensher, D. and Dalvi, M., editors, *Determinants of Travel Choice*, pages 335–357. Saxon House, Sussex.

Dawes, R. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, 25(1):1–17.

Duffy, J., Ochs, J., and Vesterlund, L. (2007). Giving little by little: Dynamic voluntary contribution games. *Journal of Public Economics*, 91(9):1708–1730.

Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H. (2006). The framing of games and the psychology of strategic play. *CeDEx Discussion Paper Series*.

Farewell, V. (1982). A note on regression analysis of ordinal data with variability of classification. *Biometrika*, 69(3):533–538.

Farrell, J. and Maskin, E. (1989). Renegotiation in repeated games. *Games and Economic Behavior*, 1(4):327–360.

Fehr, E. and Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

Fershtman, C. and Nitzan, S. (1991). Dynamic voluntary provision of public goods. *European Economic Review*, 35(5):1057–1067.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Gale, D. (2001). Monotone games with positive spillovers. *Games and Economic Behavior*, 37(2):295–320.

Güth, W., Levati, M., Sutter, M., and van der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, 91(5-6):1023–1042.

Hamilton, S. and Zilberman, D. (2006). Green markets, eco-certification, and equilibrium fraud. *Journal of Environmental Economics and Management*, 52(3):627–644.

Kübler, D. and Weizsäcker, G. (2004). Limited depth of reasoning and failure of cascade formation in the laboratory. *Review of Economic Studies*, 71(2):425–441.

Lee, B. (1999). Calling patterns and usage of residential toll service under self selecting tariffs. *Journal of Regulatory economics*, 16(1):45–82.

Levine, D. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1(3):593–622.

Lockwood, B. and Thomas, J. (2002). Gradualism and irreversibility. *Review of Economic Studies*, 69(2):339–56.

Loftin, L. (1999). Protecting the charitable investor: A rational for donor enforcement of restricted gifts. *B.U. Public Law Journal*, 8:361–585.

Marx, L. and Matthews, S. (2000). Dynamic voluntary contribution to a public project. *Review of Economic Studies*, 67(2):327–358.

Matesanz, R. (1998). Cadaveric organ donation: comparison of legislation in various countries of Europe. *Nephrology Dialysis Transplantation*, 13(7):1632–35.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2):109–142.

McFadden, D. (1978). Modelling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J., editors, *Spatial interaction theory and planning models*, pages 75–96. North Holland, Amsterdam.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470.

McKelvey, R. and Palfrey, T. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

McKelvey, R. and Palfrey, T. (1998). Quantal response equilibria for extensive form games. *Experimental Economics*, 1(1):9–41.

OTF (2008). Organs for Transplants: A Report from the Organ Donation Taskforce. *Department of Health, UK*.

Potters, J., Sefton, M., and Vesterlund, L. (2007). Leading-by-example and signaling in voluntary contribution games: An experimental study. *Economic Theory*, 33(1):169–182.

Rogers, B., Palfrey, T., and Camerer, C. (2009). Heterogeneous quantal response equilibrium and cognitive hierarchies. *Journal of Economic Theory*, 144(4):1440–1467.

Romano, R. and Yildirim, H. (2005). On the endogeneity of Cournot-Nash and Stackelberg equilibria: Games of accumulation. *Journal of Economic Theory*, 120(1):73–107.

Ross, L., Greene, D., and House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3):279–301.

Schelling, T. (1960). *The strategy of conflict*. Harvard University Press Cambridge, Massachusetts.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.

Shaw, W. and Ozog, M. (1999). Modeling overnight recreation trip choice: Application of a repeated nested multinomial logit model. *Environmental and Resource Economics*, 13(4):397–414.

Stahl, D. and Wilson, P. (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior and Organization*, 25(3):309–327.

Stahl, D. and Wilson, P. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254.

Train, K., McFadden, D., and Ben-Akiva, M. (1987). The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *The Rand Journal of Economics*, 18(1):109–123.

Weizsäcker, G. (2003). Ignoring the rationality of others: Evidence from experimental normal-form games. *Games and Economic Behavior*, 44(1):145–171.

Whitten, G. and Palmer, H. (1996). Heightening comparativists' concern for model choice: Voting behavior in Great Britain and the Netherlands. *American Journal of Political Science*, 40(1):231–260.

A   Relegated proofs and definitions

Proof of Proposition 1   Fix a subgame-perfect equilibrium σ. We claim that for all histories $h \in H^+$ and all $i \in N$,

$$\sigma_i(h)(a_i) = \begin{cases} 1, & \text{if } a_i = \min A_i(h|H^+), \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

This claim is obviously satisfied for all histories h leading to the state $\omega(h) = \max A$ (recall the definition of "states" in Footnote 8) . The following shows that, for all $a \in A$, the claim holds for all $h : \omega(h) = a$ if it holds for all $h : \omega(h) > a$. Assume the contrary and find $i \in N$ such that $\sigma_i(h)(a_i') > 0$ for some $h : \omega(h) = a$ and $a_i' > a_i$. Let $\eta \in \Delta(A)$ denote the probability distribution on the set of action profiles, according to σ in the present round after history h. Let $\eta' \in \Delta(A)$ denote the alternative probability distribution on the action space if i deviates unilaterally toward not increasing his contribution, i.e. as described in Eq. (19). Note that the actions played in the present round equate with the eventual contributions of all players (in case all $i' \in N$ stick with $\min A_{i'}(h|H^+)$ this follows from the definition of the game, and otherwise a state $\omega' > \omega$ is reached where it holds by the induction assumption). By Eq. (1), player i's expected payoff is greater under η' than under η, and hence the unilateral deviation is profitable.

To conclude, we show that since the claim holds for all histories $h : \omega(h) = \max A$ and since it holds for all $h : \omega(h) = a$ if it holds for all $h : \omega(h) > a$, for all $a \in A$, it

33

must hold for all $h \in H^+$. Define $\Omega^0 = \max A$, and iteratively, for all $t > 0$,

$$\Omega^t = \max \left(A \setminus \cup_{t' < t} \Omega^{t'} \right) \qquad \text{using } \max \Omega' = \left\{ \omega \in \Omega' \mid \nexists \omega' \in \Omega' : \omega' > \omega \right\}. \quad (20)$$

After round t of the induction, the claim is established for all histories $h : \omega(h) \in \cup_{t' \leq t} \Omega^{t'}$. Since the set A is finite, there exists $T < \infty$ such that $\cup_{t \leq T} \Omega^t = A$. $\qquad \square$

Proof of Lemma 1

1. We show that if $a' \neq a'' \in A^s$ such that neither $a' > a''$ nor $a' <''$, then $a''' = (\min\{a_i', a_i''\})_{i \in N} \in A^s$, which implies the lemma. Define $i \in N$ such that $a_i' > a_i''$; hence $a_j' < a_j''$ for $j \neq i$. By Eq. (1), this implies $p_i(a''') > p_i(a')$ and $p_j(a''') > p_j(a'')$. Since a' Pareto dominates all steady states $a \in A^s : a < a'$, this implies that i is better off in a''' than in all steady steady states $a \in A^s : a < a'$, which includes all $a \in A^s : a < a'''$. Similarly, since a'' Pareto dominates all $a \in A^s : a < a''$, j is better off in a''' than in all $a \in A^s : a < a'''$. By construction of A^s, this implies $a''' \in A^s$, and this implies a contradiction to the assumption $a', a'' \in A^s$. For, neither a' nor a'' Pareto dominate a''', although $a' > a'''$ and $a'' > a'''$ holds by construction.

2. Assume that a^* is not Pareto efficient, i.e. that there exists $a \in A$ that Pareto dominates a^*. By Lemma 1, $\max A^s$ is unique, and hence all $a' \in A^s : a' \neq a^*$ satisfy $a' < a^*$. By Eq. (9), this implies that a^* Pareto dominates all $a' \in A^s : a' \neq a^*$, and hence a Pareto dominates all $a' \in A^s$. By definition of A^s, Eq. (9), this implies $a \in A^s$, and thus contradicts either uniqueness of $\max A^s$ or that $a^* = \max A^s$. $\qquad \square$

Proof of Proposition 2 We are going to show that for all h, if h induces the state $\omega(h) = a$, then the final contributions are equal to

$$O(a) = \max\{a' \leq a \mid a' \in A^s\}, \quad (21)$$

with probability 1. By assumption Eq. (10), $O(a)$ is well-defined. This claim is trivially satisfied for all $h : \omega(h) = \min A$. The following shows that, for all $a \in A$, it

34

holds for all $h : \omega(h) = a$ if it holds for all $h : \omega(h) < a$. The proof is completed by an induction very similar to the one described in the proof of Prop. 1.

Taking $a \in A$ as given, define $a^* = O(a)$. First assume $a = a^*$. We show that for all $i \in N$, action a_i weakly dominates all $a'_i < a_i$ in response to $a'_{-i} \leq a_{-i}$. To see this, pick any $a'_i < a_i$ and $a'_{-i} \leq a_i$ of opponents' actions. On the one hand, if $a'_{-i} = a_{-i}$, then a_i is strictly more profitable than any $a'_i < a_i$. This follows by the assumption that a Pareto dominates all steady states $a'' \in A^s : a'' < a$, and hence it is also more profitable than the expected payoff from any probability distribution over steady states $a'' \leq a$ with support on steady states strictly below a. On the other hand, consider the case $a'_{-i} < a_{-i}$. Assume that some $a'_i < a_i$ is a better response to a'_{-i} than a_i. Let $o = O(a_i, a'_{-i})$ and $o' = O(a'_i, a'_{-i})$ denote the respective outcomes under σ, noting that these outcomes are deterministic and unique due to the induction assumption. Since a_i is not a best response to a'_{-i}, it follows that $p_i(o) < p_i(o')$. By uniqueness of steady states as assumed in Eq. (10), this can be satisfied only if either $o' > o$ or $o > o'$, and by the Pareto relation between alternative steady states, see Eq. (9), $p_i(o) < p_i(o')$ can be satisfied only if $o' > o$. Since $(a_i, a'_{-i}) > (a'_i, a'_{-i})$, however, this contradicts $o = O(a_i, a'_{-i})$ and $o' = O(a'_i, a'_{-i})$ based on $O(\cdot)$ as defined above. In turn, a_i weakly dominates all $a'_i < a_i$. By the definition of limiting logit equilibria, Eq. (7) for $\lambda \to \infty$, this implies that all players $i \in N$ stick to a_i after history h with probability 1.

Second assume $a \neq a^*$, which by definition implies $a^* < a$. Similar to the previous case, it can be shown that players do not move beyond a^*, i.e. with probability 1 all $i \in N$ contribute $a'_i \geq a^*_i$ in all limiting logit equilibria. Also, since $a \notin A^s$ has to be satisfied, there $\exists i \in N : p_i(a) < p_i(a^*)$. By Eq. (2), this requires $a_i > a^*_i$, and this in turn implies that a cannot be a steady state in a limiting logit equilibrium. Assume the opposite, i.e. that all $i' \in N$ stick with $a_{i'}$ after history h with positive probability. Hence, if i deviates unilaterally after h, by shifting the probability weight he puts on a_i toward a^*_i, then a^* results with probability 1 and he is better off. □

Voluntary Contributions by Consent or Dissent
Supplementary material

Jonathan H.W. Tan   Yves Breitmoser   Friedel Bolle

B   Experimental instructions

Note: In the instructions, we referred to each "game" as a "stage" of the experiment. Square brackets denote alternative text that was applied in the *D* treatments.

Stage 1 Instructions   You are participating in an experiment on decision-making. The experiment is divided into 12 stages. In each stage, the computer will randomly assign you into 4 *groups* of 3 *co-participants* per group. Participants will be re-assigned into new groups for each stage. You will be paid based on your earnings in one *winning stage*, randomly chosen by the computer.

   Task: Each stage consists of rounds. In each stage you will have 10 Euros, from which you can either keep or invest. At the start of each stage, each participant's investment is automatically set at 0 [10] Euros. In each subsequent round, you decide whether to increase [decrease] or maintain (but *not* decrease [increase]) your investment. To do so, click on a box stating your chosen investment. A stage ends if all co-participants in that group maintain their investments in the same round. The final investments at the end of the stage determine your earnings. For each Euro invested, each co-participant in your group (including you) receives X Euros. X is a factor that stays constant within a stage, but varies across stages. You will be informed of the factor for each stage on the computer screen. Generally, you earn

$$10 - \text{your investment} + X \text{ (total group investment)} \tag{22}$$

After everyone in the group has decided, you will be informed of the current investment of each co-participant, and if the stage continues or ends then. At the end of each stage, you will be informed of the final investment of each co-participant, and how much you have earned.

Example: *Take for example a factor of X = 0.5. If your final investment is 5 Euros, co-participant A's final investment is 4 Euros, and co-participant B's final investment is 6 Euros, then the total group investment is 15 Euros. The final earnings for the stage is 12.5 Euros for you, 13.5 Euros for co-participant A, and 11.5 Euros for co-participant B.*

Please fill in the questionnaire and call for the experimenter before you begin. Feel free to raise your hand for assistance at any point of the experiment.

C Extended model comparison

This section extends the analysis of strategic choice in three directions.

First, we consider the possibility that utility systematically differs from pecuniary payoffs. Formally we consider *altruism* (spite) in the Cobb-Douglas utility function

$$u_i(\pi) = \pi_i^{1-\mu} \cdot \prod_{j \neq i} \pi_j^{\mu/(n-1)} \tag{23}$$

with *n* as the number of players.[22] We chose Cobb-Douglas altruism, rather than linear altruism, because it provides a simple non-strategic explanation for our observation that player's pledges are positively correlated with their opponents' pledges ($\bar{c}$, $\underline{c}$), e.g. Eqs. (13)–(14). Note also that, in contrast to the existing literature on interdependent preferences, e.g. Levine (1998), Fehr and Schmidt (1999), and Charness and Rabin (2002), we discuss joint modeling of random utility and interdependent preferences, which allows us to evaluate their complementary relevance.

Second, Weizsäcker (2003) and more generally Rogers et al. (2009) argued that subjects tend to deviate from logit equilibrium because they are not homogenous. That is, different subjects have different precisions λ in maximizing utility and they

---

[22]Note that we also considered a utility function expression inequity aversion, but the overall fit was worse, so further discussion can be skipped.

also have different perceptions in how precise their opponents are. In choice theory, the former is known as mixed logit modeling (McFadden and Train, 2000), but following Rogers et al. (2009), we refer to it *heterogenous logit equilibrium* (HLE). Similar to Weizsäcker (2003), we assume that individual precisions are gamma-distributed. Here, a player with specific λ plays the logit equilibrium corresponding with that precision, i.e. players assume that the opponents are exactly as precise as they are themselves.[23] In this sense, the players are both heterogenous and subjective in the sense of Rogers et al. (2009). If we drop heterogeneity but maintain subjectivity, which we do alternatively, we obtain a variant of *subjective logit equilibria* (SLE) where all subjects have precision λ_1 and assign precision λ_2 to their opponents. Weizsäcker (2003) and Rogers et al. (2009) discuss many further variations of heterogeneity and subjectivity, but for our basic comparison of judgmental biases, we shall skip them.

Third, an alternative approach to model failure of equilibration in dynamic games is known from chess: the "assumption" that players do not backward induce the whole game, but consider only the moves up to T rounds into the future and evaluate the various resulting situations according to some evaluation rule. The relevance of this bias is largely unexplored in experimental analyses, the most closely related analysis considers limited depth of reasoning in cascade formation (Kübler and Weizsäcker, 2004), but the approach is well-established in computer science. We examine choice-theoretic models of *limited look-ahead* for (up to) $T = 4$ steps and two simple evaluation rules. The considered evaluation rules are that players assume everybody would randomize uniformly starting T rounds into the future (*LLA I*) or that everybody sticks with their then standing pledges T rounds into the future (*LLA II*). With respect to the T rounds that are analyzed by the players in these models, we allow that the precision parameter λ varies with the number of rounds $t \leq T$ they look into the future.[24]

---

[23]This consensus hypothesis is also known as 'false' consensus effect (e.g. Ross et al., 1977), as subjects have been found to "overestimate the degree to which they are like others" (Dawes, 1989, p. 1).

[24]Hence, there are four precision parameters $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. The current round is played with precision λ_4 by all players, the next one with λ_3, and so on.

Table 6: BICs for the extended model comparison, where $BIC = -LL + (\#Pars)/2 * ln(\#Obs)$

Model	#Pars	Blocks 1+2					Block 1					Block 2				
		All	*U*	*D*	*DR*	*DRF*	All	*U*	*D*	*DR*	*DRF*	All	*U*	*D*	*DR*	*DRF*
Logit equilibrium																
LE	1	16057	3302	4294	4522	3950	8572	1772	2332	2253	2225	7467	1510	1965	2277	1727
LE + Altr	2	15990	3209	4301	4537	3965	8552	1741	2337	2263	2233	7420	1453	1970	2283	1737
Subjective logit equilibrium																
SLE	2	15938	3189	4271	4518	3982	8526	1738	2322	2255	2233	7419	1456	1956	2272	1756
SLE + Altr	3	15935	3186	4273	4522	3987	8524	1695	2353	2272	2237	7399	1431	1965	2282	1754
Heterogenous logit equilibrium																
HLE	2	15542	3134	4191	4383	3856	8213	1695	2269	2183	2088	6688	1387	1748	2024	1550
Limited Look-Ahead																
LLA I	4	15967	3946	3999	4293	3773	8463	2040	2199	2139	2130	7344	1917	1735	2138	1598
LLA II	4	15957	3306	4265	4499	3928	8516	1790	2310	2251	2209	7273	1555	1868	2220	1673
Logit equilibrium in finite mixture with level-*k*																
LE + Lev 0	2	15831	3252	4212	4456	3934	8477	1749	2299	2242	2208	7365	1504	1932	2223	1728
LE + Lev 0–1	4	15709	3119	4220	4469	3944	8369	**1636**	2307	2258	2220	7072	1498	1834	2134	1650
LE + Lev 0–2	6	15718	3127	4228	4477	3952	8386	1644	2314	2266	2228	6998	1418	1842	2143	1660
LE + Lev 0–3	8	15493	**3068**	4223	4420	3870	8249	1656	2299	2211	2170	6910	**1351**	1850	2142	1656
LE × 2 + Lev 0	4	15345	3195	4055	4316	3823	8300	1737	2235	2193	2180	7282	1496	1923	2212	1695
LE × 2 + Lev 0–1	6	15290	3138	4063	4326	3829	8236	1667	2243	2204	2188	6986	1416	1841	2140	1655
LE × 2 + Lev 0–2	8	15277	3113	4071	4338	3842	8234	1657	2252	2214	2198	6950	1375	1849	2149	1665
Nested logit equilibrium																
NLE + Lev 0	3	**13921**	3985	**3313**	**3636**	**3021**	**7620**	1840	**1966**	**1990**	**1863**	**6136**	1971	**1357**	**1661**	**1185**
Benchmark concepts																
Naive model ($\lambda = 0$)	0	18050	5437	4180	4469	3963	9541	2815	2275	2242	2209	8509	2622	1905	2228	1755
Prop. odds logit regr.	17	12111	3917	3130	2961	2286	6898	2032	1903	1653	1480	5159	1903	1259	1326	834

D Parameter estimates

Note: We report three sets of maximum likelihood estimates per model: one for the whole data set, and one each for block 1 and block 2. In parentheses, the respective standard errors are given.

Table 7: Parameter estimates of the strategic choice models

(a) Logit models

Blocks $1+2$		Block 1		Block 2	
λ	μ	λ	μ	λ	μ
3.297 (0.048)		2.990 (0.062)		3.656 (0.077)	
2.380 (0.177)	−0.164 (0.048)	2.181 (0.147)	−0.126 (0.028)	2.664 (0.123)	−0.151 (0.022)

(b) SLE models

Blocks $1+2$			Block 1			Block 2		
λ_1	λ_2	μ	λ_1	λ_2	μ	λ_1	λ_2	μ
2.661 (0.035)	4.789 ($6e{-}05$)		2.541 ($5e{-}05$)	4.777 (0.002)		2.777 (0.012)	4.788 ($7e{-}05$)	
2.558 (0.122)	4.820 (0.120)	−0.013 (0.002)	2.588 (0.120)	1.717 (0.081)	−0.089 (0.015)	2.253 (0.057)	3.166 (0.090)	−0.186 (0.004)

(c) Heterogenous logit models

Blocks $1+2$		Block 1		Block 2	
Shape	Scale	Shape	Scale	Shape	Scale
0.683 (0.055)	5.967 (0.100)	0.788 (0.001)	7.301 (2.644)	0.693 (0.051)	15.237 (0.170)

Table 7: Parameter estimates (continued)

(d) Limited look-ahead models

	LLA I				LLA II			
	λ_4	λ_3	λ_2	λ_1	λ_4	λ_3	λ_2	λ_1
Overall	4.410 (0.001)	4.313 (0.281)	9.758 (0.701)	8.921 (0.003)	2.583 (0.064)	8.278 (0.580)	3.850 (0.253)	0.404 (0.008)
Block 1	3.868 (0.001)	4.691 (0.001)	6.344 (0.001)	6.902 (0.001)	2.126 (0.069)	12.856 (1.210)	4.102 (0.001)	0.416 (0.009)
Block 2	4.043 (0.002)	2.905 (0.004)	6.643 (0.001)	11.460 (0.010)	3.763 (0.001)	34.315 (2.838)	3.306 (0.207)	0.322 (0.011)

(e) One logit equilibrium type

ρ_0	Logit Eq.		Level 1		Level 2		Level 3	
	ρ_E	λ_E	ρ_1	λ_1	ρ_2	λ_2	ρ_3	λ_3
Overall								
.423	0.577 (0.045)	3.604 (0.060)						
.468	0.412 (0.049)	3.289 (0.079)	0.119 (0.027)	4.815 (0.240)				
.468	0.412 (0.049)	3.289 (0.079)	0.119 (0.027)	4.815 (0.240)	0 (−)	4.818 (1.303)		
.208	0.092 (0.024)	2.956 (0.096)	0.031 (0.012)	8.339 (0.094)	0.154 (0.037)	28.212 (2.551)	0.513 (0.047)	2.622 (0.106)
Block 1								
.370	0.630 (0.053)	3.293 (0.073)						
.470	0.370 (0.060)	2.514 (0.085)	0.159 (0.031)	4.702 (0.214)				
.470	0.370 (0.060)	2.514 (0.086)	0.159 (0.031)	4.702 (0.214)	0 (−)	4.708 (0.770)		
.184	0.196 (0.037)	2.468 (0.073)	0.015 (0.008)	8.228 (0.037)	0.144 (0.032)	64.525 (0.000)	0.459 (0.048)	3.973 (0.277)
Block 2								
.293	0.706 (0.050)	4.546 (0.013)						
.315	0.445 (0.047)	4.903 (0.122)	0.240 (0.036)	3.575 (0.106)				
.316	0.390 (0.045)	4.913 (0.160)	0.230 (0.036)	3.825 (0.119)	0.063 (0.024)	6.136 (0.322)		
.149	0.321 (0.022)	4.948 (1.084)	0.073 (0.022)	5.283 (0.313)	0.062 (0.024)	8.435 (0.700)	0.392 (0.038)	2.175 (0.123)

Table 7: Parameter estimates (continued)

(f) Two logit equilibrium types

	Logit Eq. 1		Logit Eq. 2		Level 1		Level 2	
ρ_0	ρ_{E_1}	λ_{E_1}	ρ_{E_2}	λ_{E_2}	ρ_1	λ_1	ρ_2	λ_2
Overall								
.228	0.377 (0.041)	3.482 (0.060)	0.394 (0.037)	4.913 (0.071)				
.261	0.246 (0.037)	3.478 (0.130)	0.378 (0.036)	4.933 (0.206)	0.113 (0.022)	4.110 (0.300)		
.281	0.121 (0.034)	2.696 (0.086)	0.377 (0.037)	4.914 (0.073)	0.182 (0.033)	3.675 (0.086)	0.039 (0.017)	4.741 (0.200)
Block 1								
.253	0.375 (0.050)	3.007 (0.087)	0.372 (0.047)	4.913 (0.043)				
.286	0.258 (0.048)	2.970 (0.119)	0.349 (0.046)	4.903 (0.069)	0.106 (0.026)	4.626 (0.234)		
.254	0.222 (0.048)	2.452 (0.079)	0.334 (0.056)	4.904 (0.066)	0.104 (0.026)	4.610 (0.227)	0.087 (0.092)	5.010 (0.970)
Block 2								
.215	0.421 (0.053)	3.772 (0.113)	0.363 (0.049)	6.160 (0.001)				
.238	0.313 (0.049)	3.804 (0.175)	0.378 (0.044)	4.947 (0.236)	0.069 (0.021)	6.508 (0.698)		
.231	0.286 (0.047)	3.769 (0.172)	0.351 (0.043)	4.948 (0.289)	0.071 (0.022)	5.777 (0.359)	0.060 (0.027)	10.280 (0.941)

(g) NLE models

Blocks $1+2$			Block 1			Block 2		
λ	κ	ρ	λ	κ	ρ	λ	κ	ρ
3.635 (0.002)	0.243 (0.001)	0.976 (0.009)	4.249 ($1e-4$)	0.374 ($7e-05$)	0.892 (0.009)	3.683 ($4e-05$)	0.228 ($1e-05$)	0.966 (0.015)

Table 8: Proportional odds regression of final contributions

	Blocks $1+2$		Block 1		Blocks $1+2$	
	Model 1	Model 1'	Model 2	Model 2'	Model 3	Model 3'
Coefficients						
α	4.157	4.085	6.058	5.953	2.447	2.457
	(0.368)	(0.367)	(0.532)	(0.530)	(0.517)	(0.516)
D	-0.952	0.593	-1.508	-0.108	0.265	1.231
	(0.201)	(0.098)	(0.312)	(0.139)	(0.768)	(0.140)
R	-0.229	-0.526	-0.141	-0.486	-0.272	-0.633
	(0.216)	(0.102)	(0.333)	(0.148)	(0.808)	(0.144)
RF	-0.356	0.131	-0.842	0.019	-0.507	0.275
	(0.218)	(0.102)	(0.344)	(0.145)	(0.847)	(0.149)
G	-0.046	0.132	-0.099	0.222	-0.034	0.044
	(0.019)	(0.010)	(0.057)	(0.030)	(0.055)	(0.030)
$D \times G$	0.246		0.411		0.102	
	(0.028)		(0.082)		(0.080)	
$R \times G$	-0.050		-0.105		-0.038	
	(0.030)		(0.086)		(0.084)	
$RF \times G$	0.082		0.234		0.084	
	(0.031)		(0.088)		(0.089)	
Intercepts						
0|1	1.094	2.230	1.723	2.804	0.434	1.193
	(0.232)	(0.210)	(0.342)	(0.304)	(0.592)	(0.399)
1|2	1.407	2.533	2.067	3.139	0.722	1.481
	(0.232)	(0.211)	(0.344)	(0.305)	(0.592)	(0.399)
2|3	1.729	2.843	2.462	3.523	0.975	1.734
	(0.233)	(0.212)	(0.345)	(0.308)	(0.592)	(0.400)
3|4	2.045	3.144	2.875	3.917	1.202	1.960
	(0.233)	(0.214)	(0.346)	(0.311)	(0.593)	(0.400)
4|5	2.290	3.380	3.136	4.166	1.441	2.199
	(0.234)	(0.215)	(0.348)	(0.313)	(0.593)	(0.401)
5|6	2.544	3.623	3.441	4.461	1.654	2.412
	(0.235)	(0.217)	(0.350)	(0.316)	(0.594)	(0.402)
6|7	2.749	3.821	3.695	4.709	1.822	2.579
	(0.236)	(0.218)	(0.352)	(0.319)	(0.594)	(0.403)
7|8	2.927	3.991	3.915	4.924	1.972	2.728
	(0.237)	(0.220)	(0.354)	(0.322)	(0.594)	(0.404)
8|9	3.200	4.252	4.275	5.274	2.191	2.945
	(0.239)	(0.222)	(0.358)	(0.326)	(0.595)	(0.406)
9|10	3.461	4.501	4.488	5.482	2.482	3.234
	(0.240)	(0.224)	(0.360)	(0.329)	(0.596)	(0.408)
BIC	4927	4981	2496	2514	2440	2432

Table 9: Proportional odds estimates of strategies

	Aggregate data set			D treatments			U treatments		
	Bl. $1+2$	Block 1	Block 2	Bl. $1+2$	Block 1	Block 2	Bl. $1+2$	Block 1	Block 2
$\underline{c}$	0.244 (0.011)	0.126 (0.014)	0.346 (0.017)	0.147 (0.012)	0.090 (0.016)	0.256 (0.018)	0.472 (0.030)	0.081 (0.042)	0.725 (0.049)
$\bar{c}$	0.149 (0.012)	0.187 (0.015)	0.145 (0.018)	0.236 (0.014)	0.255 (0.019)	0.207 (0.022)	0.047 (0.024)	0.153 (0.031)	0.033 (0.038)
α	0.601 (0.219)	1.108 (0.292)	1.166 (0.350)	3.06 (0.271)	4.321 (0.358)	0.092 (0.428)	1.235 (0.399)	0.418 (0.545)	3.242 (0.662)
D	0.624 (0.067)	0.623 (0.087)	0.686 (0.104)						
R	−0.171 (0.063)	−0.138 (0.082)	−0.217 (0.098)	−0.176 (0.062)	−0.322 (0.083)	−0.347 (0.096)			
RF	−0.474 (0.067)	−0.026 (0.085)	−0.630 (0.111)	−0.291 (0.067)	−0.039 (0.086)	−0.584 (0.111)			
G	0.054 (0.006)	0.109 (0.017)	−0.089 (0.019)	0.089 (0.008)	0.242 (0.021)	0.074 (0.024)	0.012 (0.011)	−0.029 (0.032)	0.002 (0.032)
0|1	0.504 (0.128)	0.922 (0.174)	−0.117 (0.279)	1.832 (0.161)	2.610 (0.216)	0.74 (0.356)	0.247 (0.212)	−0.365 (0.313)	1.541 (0.484)
1|2	1.168 (0.128)	1.559 (0.174)	0.596 (0.279)	2.323 (0.162)	3.175 (0.217)	1.187 (0.355)	1.312 (0.214)	0.497 (0.311)	2.551 (0.493)
2|3	1.821 (0.129)	2.156 (0.175)	1.197 (0.280)	2.883 (0.164)	3.813 (0.221)	1.624 (0.356)	2.061 (0.217)	1.111 (0.313)	3.499 (0.501)
3|4	2.402 (0.131)	2.710 (0.177)	1.781 (0.281)	3.368 (0.167)	4.356 (0.225)	1.998 (0.357)	2.770 (0.221)	1.789 (0.316)	4.279 (0.507)
4|5	2.883 (0.133)	3.156 (0.179)	2.289 (0.283)	3.801 (0.169)	4.798 (0.229)	2.470 (0.36))	3.328 (0.225)	2.330 (0.319)	4.863 (0.511)
5|6	3.418 (0.135)	3.640 (0.182)	2.918 (0.287)	4.320 (0.173)	5.331 (0.234)	3.070 (0.364)	3.871 (0.229)	2.853 (0.324)	5.650 (0.519)
6|7	3.799 (0.137)	4.026 (0.184)	3.346 (0.289)	4.667 (0.175)	5.733 (0.237)	3.405 (0.366)	4.410 (0.236)	3.258 (0.329)	6.343 (0.528)
7|8	4.175 (0.139)	4.385 (0.187)	3.750 (0.292)	4.972 (0.177)	6.078 (0.240)	3.715 (0.369)	5.022 (0.245)	3.710 (0.336)	7.163 (0.543)
8|9	4.632 (0.141)	4.860 (0.190)	4.277 (0.295)	5.436 (0.18))	6.535 (0.244)	4.221 (0.373)	5.826 (0.261)	4.397 (0.355)	8.279 (0.566)
9|10	5.102 (0.144)	5.265 (0.194)	4.780 (0.298)	5.830 (0.182)	6.859 (0.246)	4.696 (0.376)	6.959 (0.294)	5.221 (0.393)	9.947 (0.607)
BIC	12111	6898	5159	8188	4912	3265	3795	1981	1781

Table 10: Nonparametric tests of mean contributions and p-values (in parentheses) for $\alpha = 0.4$ (in cells below diagonal) and $\alpha = 0.6$ (in cells above diagonal)

Block 1	U	D	DR	DRF
U	–	−1.599 (0.1099)	0 (1)	−1.155 (0.2482)
D	1.47 (0.1416)	–	−0.735 (0.4624)	−0.735 (0.4624)
DR	1.732 (0.0833)	0 (1)	–	−0.289 (0.7728)
DRF	0 (1)	−1.470 (0.1416)	−1.443 (0.1489)	–

Block 2	U	D	DR	DRF
U	–	−0.735 (0.4624)	−1.155 (0.2482)	−1.155 (0.2482)
D	−2.449 (0.0143)	–	0.245 (0.8065)	0.245 (0.8065)
DR	−1.597 (0.1102)	0.980 (0.3272)	–	−1.443 (0.1489)
DRF	−2.309 (0.0209)	−0.490 (0.6242)	−1.016 (0.3094)	–

Game 1	U	D	DR	DRF
U	–	1.715 (0.0864)	1.452 (0.1465)	2.309 (0.0209)
D	2.460 (0.0139)	–	1.230 (0.2187)	1.715 (0.0864)
DR	2.323 (0.0202)	1.235 (0.2168)	–	0.871 (0.3836)
DRF	2.191 (0.0284)	−0.861 (0.3893)	−0.877 (0.3807)	–

Game 7	U	D	DR	DRF
U	–	−0.246 (0.8057)	−0.289 (0.7728)	0 (1)
D	−2.460 (0.0139)	–	0.371 (0.7110)	0.738 (0.4606)
DR	−1.732 (0.0833)	1.722 (0.0851)	–	0.289 (0.7728)
DRF	−1.888 (0.0591)	−0.246 (0.8057)	−1.155 (0.2482)	–

Game 6	U	D	DR	DRF
U	–	−1.960 (0.0500)	−1.016 (0.3904)	−2.033 (0.0421)
D	0.738 (0.4606)	–	0.980 (0.3272)	−0.249 (0.8033)
DR	0.290 (0.7715)	0.256 (0.7981)	–	−1.162 (0.2454)
DRF	−1.155 (0.2482)	−0.984 (0.3252)	0.581 (0.5614)	–

Game 12	U	D	DR	DRF
U	–	−1.722 (0.0851)	−1.016 (0.3094)	−1.607 (0.1081)
D	−2.460 (0.0139)	–	0.984 (0.3252)	−0.507 (0.6121)
DR	−2.323 (0.0202)	0.494 (0.6213)	–	−0.871 (0.3836)
DRF	−1.742 (0.0814)	0.761 (0.4469)	0 (1)	–