

Nax, Heinrich H.; Murphy, Ryan O.; Duca, Stefano; Helbing, Dirk

## Article

# Contribution-based grouping under noise

Games

## Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

*Suggested Citation:* Nax, Heinrich H.; Murphy, Ryan O.; Duca, Stefano; Helbing, Dirk (2017) : Contribution-based grouping under noise, Games, ISSN 2073-4336, MDPI, Basel, Vol. 8, Iss. 4, pp. 1-23,  
<https://doi.org/10.3390/g8040050>

This Version is available at:

<https://hdl.handle.net/10419/179160>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

# Contribution-Based Grouping under Noise

Heinrich H. Nax <sup>1,\*</sup>, Ryan O. Murphy <sup>2,3</sup>, Stefano Duca <sup>1</sup> and Dirk Helbing <sup>1</sup>

<sup>1</sup> Computational Social Science, ETH Zürich, 8092 Zürich, Switzerland; sduca@ethz.ch (S.D.); dirk.helbing@gess.ethz.ch (D.H.)

<sup>2</sup> Department of Economics, University of Zurich, 8006 Zürich, Switzerland; ryan.murphy@econ.uzh.ch

<sup>3</sup> Morningstar Inc., Behavioral Science Group, Chicago, IL 60602, USA

\* Correspondence: hnax@ethz.ch; Tel.: +41-44-63-20594

Received: 12 October 2017; Accepted: 3 November 2017; Published: 17 November 2017

**Abstract:** Many real-world mechanisms are “noisy” or “fuzzy”, that is the institutions in place to implement them operate with non-negligible degrees of imprecision and error. This observation raises the more general question of whether mechanisms that work in theory are also robust to more realistic assumptions such as noise. In this paper, in the context of voluntary contribution games, we focus on a mechanism known as “contribution-based competitive grouping”. First, we analyze how the mechanism works under noise and what happens when other assumptions such as population homogeneity are relaxed. Second, we investigate the welfare properties of the mechanism, interpreting noise as a policy instrument, and we use logit dynamic simulations to formulate mechanism design recommendations.

**Keywords:** voluntary contributions; behavioral economics; noise; heterogeneity; mechanism design; welfare; efficiency; equality

**JEL Classification:** C73; D02; D03; D63

## 1. Motivation

Typically, individual decisions in social-dilemma interactions are not perfectly observable in the real world. Applied mechanism designers should keep this in mind when implementing mechanisms, in particular in the context of interactions that have the strategic nature of voluntary contribution games where imperfect observability is ubiquitous.<sup>1</sup> Real-world institutions are usually “noisy” or “fuzzy”, operating with non-negligible degree of imprecision. By contrast, theory investigations of mechanisms for the most part study perfect mechanisms. In this paper, as a first step towards performing robustness checks of mechanisms under relaxed assumptions more generally, we investigate the mechanism of contribution-based competitive grouping (as introduced by [5]): we relax some assumptions and investigate what happens when there is (i) noise, (ii) heterogeneity and (iii) different action spaces.

The paper is structured as follows. Next, we introduce our model. In Section 3, we first analyze how the existence of Nash equilibria depends on (iii) the strategy space of the game and on (ii) the underlying population homogeneity/heterogeneity in terms of contribution budgets. Then, we assess their robustness when we allow more and more (i) monitoring noise. In Section 4, we investigate how a social planner, interpreting monitoring noise and/or the other model characteristics as policy instruments, would trade-off efficiency and equality to maximize social welfare. Finally, we use agent-based simulations to study the logit dynamics of the game in order to quantify the effect of noisy monitoring.

<sup>1</sup> For other social-dilemma contexts, see, for example, [1,2] for imperfect public monitoring, [3] for noisy prisoners' dilemmas and [4] for team-production games with group-level information.

Our results are summarized as follows. In terms of the existence of Nash equilibria, the zero-contribution outcome is always a Nash equilibrium and becomes the unique one under too much noise. High-contribution equilibria exist if three things come together: the rate of return of the underlying contribution game is high enough, grouping is sufficiently precise, and agent homogeneity is established (either in terms of budgets or via the strategy space). In terms of welfare, implementations with an intermediate level of noise/fuzziness, enabling high efficiency gains at low inequality costs, maximize welfare in most cases. The exception is the case of heterogeneous budgets and binary (all-or-nothing) contribution decisions, where less noise is unambiguously preferable. Finally, our logit-dynamics simulations indicate that high-contribution equilibria, when existent, tend to be more robust than the zero-contribution equilibrium, and we are able to identify an optimal level of mechanism noise/fuzziness as a function of the behavioral noise that is inherent under the logit-choice rule.

## 2. The Mechanism

### 2.1. The Model

Population  $N = \{1, 2, \dots, n\}$  plays the following three-step game under common knowledge.

Step 1. Voluntary contributions:

Action Space 1. Binary: Each  $i \in N$  simultaneously decides whether to set  $c'_i = 0$  (i.e., to free-ride) or to set  $c'_i = 1$  (i.e., to contribute).

Action Space 2. Continuous: Each  $i \in N$  simultaneously decides on a proportional contribution  $c'_i \in [0, 1]$  between free-ride ( $c'_i = 0$ ) and contribute ( $c'_i = 1$ ).

Each  $i \in N$  has a budget of  $B_i \in \mathcal{R}^+$ , and his/her above action results in an effective contribution of  $c_i = c'_i \cdot B_i$ . The effective contributions yield a population contribution vector  $\mathbf{c} = \{c_i\}_{i \in N}$ . Denote by  $\mathbf{c}_{-i}$  the vector excluding  $i$ .

Budget 1. Homogeneity:  $B_i = 1$  for all  $i$ .

Budget 2. Heterogeneity:  $B_i \neq B_j \forall i \neq j$  with (w.l.o.g.)  $B_i < B_{i-1}$ . We place a very mild constraint on how different endowments can be by imposing that  $\exists X \in \mathcal{R}$  s.t.  $B_i > B_{i-1} + X \forall i$  and any  $X > 0$ .

Step 2. Fuzzy grouping based on a ‘noisy’ ranking of contributions:

Step 2.1. From ‘noisy’ rankings...: Instead of  $c_i$ , the authority observes  $x_i = c_i + e_i$ , where  $e_i$  is some i.i.d. white noise with mean zero and variance  $\sigma^2 = \frac{1-\beta}{\beta}$  where  $\beta \in (0, 1]$ . Based on the vector of observed contributions  $\mathbf{x}$ , players are ranked from highest to lowest.

$\beta$  takes the role of a “meritocracy” parameter in our setting: (i) under no meritocracy (when  $\beta \rightarrow 0$ ), all rankings are equally likely, and all players have the same expected rank<sup>2</sup>; (ii) under full meritocracy (when  $\beta = 1$ ), only “perfect” rankings are possible, that is, players who contributed more have a higher rank than players who contributed less, and ties are randomly broken; (iii) under intermediate meritocracy (when  $\beta \in (0, 1)$ ), all rankings have positive probability, but on aggregate, depending on  $\beta$ , players who contributed more have a higher expected rank than players who contributed less. As  $\beta$  increases, we move continuously from (i) to (ii).

Step 2.2. ...to fuzzy groupings: Groups form based on the noisy ranking such that  $g$  groups  $\{S_1, S_2, \dots, S_g\}$  of a fixed size  $s < n$  form. The result is a partition  $\rho$  of  $N$  (where  $s = n/g > 1$  for some  $s, g \in \mathcal{N}^+$ ) with groups  $S_p \in \rho$  (s.t.  $p = 1, 2, \dots, g$ ) where the  $s$  highest-ranked players form Group 1, the  $s$  second-highest players form Group 2, etc.<sup>3</sup>

<sup>2</sup> With a slight abuse of notation, from now on, we will write  $\beta \in [0, 1]$  and  $\beta = 0$  to indicate the “no meritocracy” case where  $\sigma \rightarrow \infty$ , and the ranking, and consequentially, the matching, of the players is (uniformly) randomly selected.

<sup>3</sup> Meritocracy  $\beta$  guides smoothly from (i) no meritocracy (grouping is random as in [6]) to (ii) full meritocracy (the case of perfect contribution-based grouping as in [5]). Note that many public goods experiments use variants of Andreoni’s random (re-)matching implementation (e.g., [6–16]); see [17,18] for reviews.

(See Appendix B for implementation examples of fuzzy grouping.)

### Step 3. Payoffs:

Step 3.1. Realized payoffs (ex post): Given  $\mathbf{c}$  and  $\rho$ , total payoffs generated in each  $S \in \rho$  are  $s + (r - 1) \sum_{i \in S} c_i$ , where  $r$  is the rate of return. Each  $i \in N$  s.t.  $i \in S \in \rho$  receives an individual payoff of:

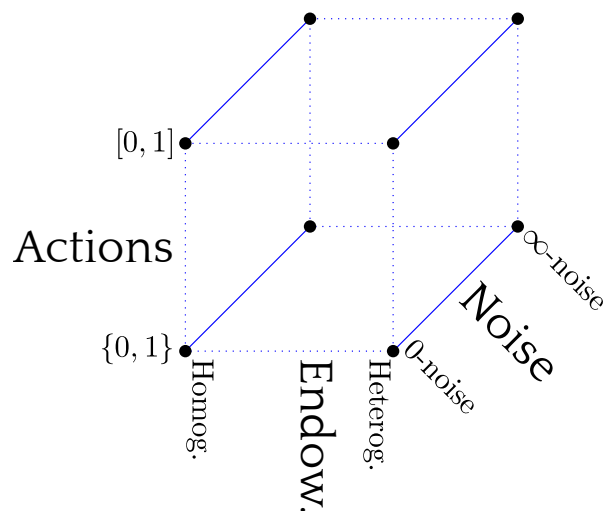
$$\phi_i(c_i | c_{-i}, \rho) = \underbrace{(1 - c_i)}_{\text{remainder from budget}} + \underbrace{(R) * \sum_{j \in S} c_j}_{\text{return from the public good}}, \quad (1)$$

where  $R := \frac{r}{s}$  is the marginal per capita rate of return; it is standard to assume a ‘social dilemma’ character of the game by setting  $R \in (\frac{1}{s}, 1)$ .<sup>4</sup> Let  $\phi = \{\phi_i\}_{i \in N}$  denote the payoff vector.

Step 3.2. Expected payoffs (ex ante): Groups form in Step 2, and payoffs materialize in Step 3, both based on the sunk contribution decisions taken during Step 1. From expression (1),  $i$ ’s expected payoff of contributing  $c_i$ , as evaluated during Step 1 given  $c_{-i}$ , is therefore:

$$\underbrace{\mathbf{E}[\phi_i(c_i | c_{-i})]}_{\text{exp. return from } c_i} = \underbrace{1}_{(i) \text{ budget}} - \underbrace{(1 - R) * c_i}_{(ii) \text{ sure loss on own contribution}} + \underbrace{R * \mathbf{E}\left[\left(\sum_{j \neq i: j \in S} c_j\right) | c_i\right]}_{(iii) \text{ exp.return from others in group}}, \quad (2)$$

Note that Term (iii), the expected return from others’ contributions, may depend on one’s own contribution if  $\beta > 0$ .



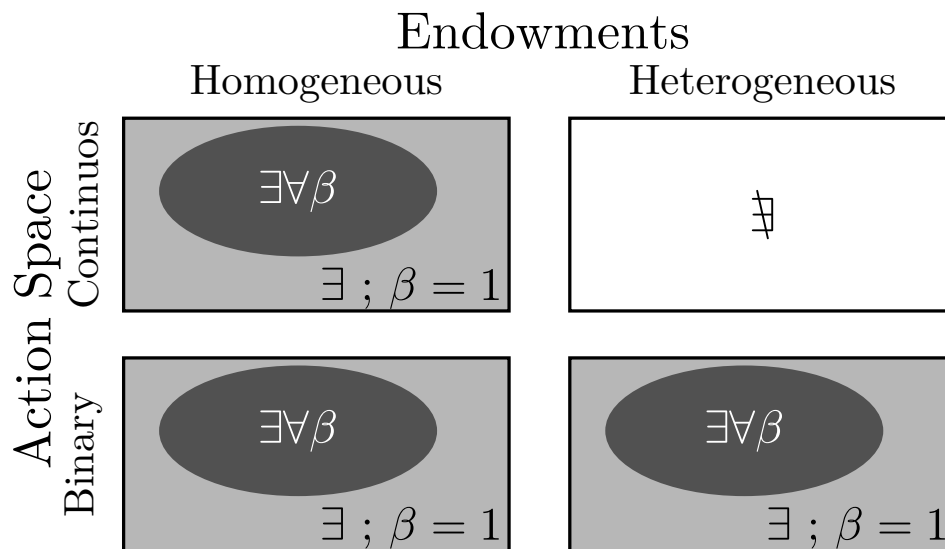
**Figure 1.** The figure illustrates the three dimensions that we vary to assess the robustness of the mechanism. Two are binary (indicated by the blue dotted lines): For one, players have either the same or different initial endowments; further, they either choose a proportional or an all-or-nothing contribution. One dimension is varied continuously (indicated by the blue continuous line): contributions are observed with different degrees of noise.

### 3. Nash Equilibria

Next, we analyze the Nash equilibria in terms of the ex ante decisions made in the games that we detailed in the previous section, where various games result from the different combinations of the

<sup>4</sup> Thus, full-contribution is collectively efficient, and zero-contribution, despite collectively inefficient, is the unique Nash equilibrium under no meritocracy.

chosen model ingredients. Figures 1 and 2 summarize the models we shall consider and the results we obtain.



**Figure 2.** The figure summarizes the existence of the “high-efficiency” equilibria described in this section. For homogeneous endowments, regardless of the action space, there exist high contribution equilibria for high enough marginal rates of return. In the case of heterogeneous endowments, only a binary action space allows for contributive equilibria. Whenever high contribution equilibria exist under full meritocracy, there exists a marginal rate of return (higher than the one for  $\beta = 1$ ) such that the high equilibria exist also with a fuzzy mechanism.

### 3.1. Game 1: ‘Baseline’

The following equilibrium structure results in the no-noise implementation with agent homogeneity and continuous action space (as in [5]). For this case, there always exists a Nash equilibrium with zero contributions; as in standard voluntary contribution games with a single group or with random grouping. We call this equilibrium the “non-contribution equilibrium”. However, non-contribution is not a dominant strategy, as it would be if groups were randomly assigned, because players may prefer to contribute when this gets them into high-contribution groups that promise higher payoffs. As a result, provided the marginal per capita rate of return  $R$  is high enough, [5] prove that there exist new asymmetric pure-strategy Nash equilibria where most players contribute and very few players free-ride.<sup>5</sup> We call these equilibria “high-contribution equilibria”. We call the threshold for which these equilibria exist  $mpcr$ ; note that this threshold is increasing in the relative group size  $s/n$ , which implies that it is, ceteris paribus, harder (easier) to satisfy for relatively large groups (populations)

High-contribution equilibria are particularly interesting, because several recent experimental studies provide support for them [5,19,20].<sup>6</sup>

<sup>5</sup> See [5], Theorem 1.

<sup>6</sup> In fact, recent evidence suggests that players may endogenously implement variants of contribution-based competitive grouping over time and then converge to the high-contribution equilibria [21].

### 3.2. Game 2: ‘Heterogeneity Extension’ (Extension 1)

#### 3.2.1. Heterogeneity and Continuous Action Space

The first negative result is that under heterogeneity, non-contribution is the only existing equilibrium. A key property of high-contribution equilibria in the baseline case is that there exists a mixed group where both defectors and contributors are placed. In these equilibria, a contributor has a high-enough probability to be placed in a group filled only with contributors and a low enough probability to be placed in the mixed group so that it is, in expectation, not beneficial to free-ride and be placed with certainty in the low group instead.

However, in [22], it is proven that, under heterogeneity, there cannot exist an equilibrium where two or more players contribute the same amount. The reason is that, if two or more players contribute the same amount, then the one with the highest endowment would have a profitable deviation in contributing slightly more due to the existence of the mixed group: contributing only slightly more, and he/she is guaranteed not to be placed in a worse group. Conversely, if players in a higher group already contribute everything, then the wealthiest player could instead contribute slightly less and still be guaranteed to remain in the same group. Therefore, when players have access to a continuous action space and can thus reduce or increase their contributions by an infinitesimal amount, then there can be no high contribution equilibrium for heterogeneous endowments.

It is important to note that this result relies on players all having different endowments. Indeed, if players were to belong only to two endowment classes, Gunnthorsdottir et al. [23] prove that there might exist an equilibrium where every player contributes everything. This equilibrium, however, requires conditions that become harder to satisfy the more classes of endowments are introduced, eventually becoming impossible if there are as many levels of endowment as there are players. Interestingly, Gunnthorsdottir et al. [19,23] also find empirical support for the high-efficiency equilibria by showing that, when existing, near-efficient equilibria are re-preferred over the worst ones.

#### 3.2.2. Binary Action Space

With a binary action space, the above argument about infinitesimal changes in contributions does not hold any more. Indeed, it can be shown (see Proposition A5 in Appendix A) that in the case of heterogeneous endowments and binary action space, there exists a threshold for the marginal rate of return such that contributive equilibria exist.

For this configuration, a maximum of  $g$  pure strategy Nash equilibria can coexist, depending on the value of  $r$ . Their structure is so that the  $k \cdot s$  ( $0 \leq k < g$ ) poorest players defect and the  $(g - k) \cdot s$  richest players contribute. The non-contribution equilibrium ( $k = 0$ ) always exists, regardless of the value of the marginal rate of return. Interestingly, for the equilibrium with  $k = 1$  to exist, the marginal rate of return has to be higher than for the equilibrium with  $k = 2$  to exist, and so on, until the lowest threshold for  $R$  such that the only contributive equilibrium is the one where all but one group is filled with contributors<sup>7</sup>.

As a corollary of Proposition A5, there also exist high-contribution equilibria for homogeneous endowments and binary action space. In this case, all thresholds collapse and reduce to the same value as in the baseline case, and all the existing equilibria are such that there exists a mixed group of contributors and defectors, in analogy with the continuous action space case.

### 3.3. Game 3: ‘Noise Extension’ (Extension 2)

In the case of a noisy mechanism, the existence of Nash equilibria summarizes as follows.

<sup>7</sup> For a marginal per capita rate of return  $R > s$ , the game is not a social dilemma anymore: “cooperate” becomes a dominant strategy, and the only existing equilibrium is for everybody to contribute everything.

Whenever the chosen model allows for high-contribution equilibria in perfect meritocracy, then there exists a noise/fuzziness level below which the same equilibria continue to exist: given any  $R > mpcr$ , there exists a  $\beta < 1$  such that the same high-contribution equilibria (possibly with more free-riders than under  $\beta = 1$ , but  $< s$ ) continue to exist as when  $\beta = 1$  (see Appendix A, Proposition A8). The minimum level of  $\beta$ , denoted by  $\beta$ , for which high-contribution equilibria exist, is an implicit function that is decreasing in  $R$  and increasing  $s/n$ .

Universal non-contribution is a Nash equilibrium for any  $\beta$  (see Appendix A, Proposition A7). The reason for this is that the only incentive for an individual to contribute a positive amount under contribution-based grouping is to be grouped with others doing likewise. When no one else contributes, there is no such incentive.

### 3.4. Remark: Mixed-Strategy Nash Equilibria

With heterogeneous endowments, symmetric mixed-strategy NE do not exist, either because the only existing equilibrium is non-contribution by all or because of the structure of the high-efficient equilibria (see Proposition A5 in Appendix A and Theorem 2 in [22]).

Similarly, for heterogeneous endowments and continuous action space, mixed-strategy equilibria do not exist (see Appendix A, Corollary A14). When contribution decisions are restricted to coarse or binary action spaces, however, it is the case that, for every  $\beta$ , there exists a  $R \in (mpcr, 1)$  at which there exists one and above which there exist two mixed-strategy Nash equilibria of the form ‘contribute fully with  $p$  and free-ride with  $1 - p$ ’, one with a high contribution probability  $\bar{p}$  and one with a low  $p$  (see Appendix A, Proposition A10).<sup>8</sup>

## 4. Welfare Comparison

You can’t have your cake and eat it, too, is a good candidate for the fundamental theorem of economic analysis. We can’t have our cake of market efficiency and share it equally. [24], p. 1.

We now turn to the point of view of a social planner and analyze the welfare properties of equilibria in terms of realized payoffs  $\phi$ .<sup>9</sup> Ex post, in contrast to ex ante, it is not always the case that high-contribution equilibria are payoff-dominant.<sup>10</sup> Hence, it is not obvious which equilibria a social planner who cares about equality will prefer. To evaluate welfare, we therefore use a variant of the social welfare function of [28], which has the advantage that, like the parameter  $\beta$  governing meritocracy, a single parameter characterizes a continuous range of social planner preferences, spanning preferences across the entire equality-efficiency tradeoff.

Social welfare: Given outcome  $(\rho, \phi)$ , let  $W_e(\phi)$  be the social welfare function measuring its welfare given the inequality aversion parameter  $e \in [0, \infty)$ :

$$W_e(\phi) = \frac{1}{n(1-e)} \sum_{i \in N} \phi_i^{1-e} \quad (3)$$

(For  $e = 1$ , expression (3) is defined by  $W_1(\phi) = \frac{1}{n} \prod_{i \in N} \phi_i$ , i.e., by the Nash product.) Expression (3) nests both the utilitarian (Bentham) and Rawlsian social welfare functions as special cases.<sup>11</sup> When  $e = 0$ , Expression (3) reduces to  $W_0(\phi) = \frac{1}{n} \sum_{i \in N} \phi_i$ , i.e., the utilitarian criterion measuring the state’s ‘efficiency’ (total payoffs) only. When  $e \rightarrow \infty$ , Expression (3) approaches  $W_\infty(\phi) = \min(\phi_i)$ , i.e., the Rawlsian criterion measuring the outcome’s welfare by the utility of

<sup>8</sup> Note that, whenever pure strategy highly-efficient equilibria exist, there could also exist several asymmetric Nash equilibria whose characterization is not easily obtained.

<sup>9</sup> Note that Harsanyi’s social welfare approach [25], by contrast, would consider ex ante payoffs, that is expected payoffs. His social welfare function is  $W_H(\phi) = \frac{1}{n} \sum_{i \in N} E[\phi_i]$ . See, for example, [26] for a discussion of ex ante versus ex post approaches.

<sup>10</sup> The work in [27] defines that outcome  $\phi$  payoff-dominates  $\phi'$  if  $\phi_i \geq \phi'_i$  for all  $i$ , and there exists a  $j$  such that  $\phi_j > \phi'_j$ .

<sup>11</sup> See, for example, [29] for a discussion of this generalization.



the worst-off. Obviously, a utilitarian social planner prefers the high-contribution equilibria to the non-contribution equilibrium, because it is more efficient. A Rawlsian, however, would prefer the non-contribution equilibrium (with perfect equality of payoffs) if any player receives a lesser payoff in the high-contribution equilibria.

Which equilibrium is preferable in terms of social welfare for any given social welfare function depends on the game considered and on the social planner's relative weights on efficiency and equality. Critical for this assessment is the inequality aversion  $e$ .

**Welfare criterion:** The social planner acts in order to maximize  $E[W_e(\phi)]$ , where  $\phi$  are evaluated according to the equilibrium selection criterion. Suppose that the social planner can set some of the rules of the game to achieve the above goal: Which conditions maximize the social welfare? Depending on the game, what is the value of  $\beta \in [0, 1]$  that maximizes  $E[W_e(\phi)]$ ?

#### 4.1. Homogeneous Endowments

In case of homogeneous endowments, there is a clear trade-off between efficiency and equality: the high gains in efficiency are obtained at the expense of the cooperative players placed in the group containing some defectors. These few players are worse off than in the case where everybody defects. A very inequality averse social planner might choose to increase the fuzziness in the observations in order to revert to the non-contribution equilibrium and thus improve the ex-post payoff of the worst off.

For an example, consider the economy illustrated in Table 1 (with  $n = 16$ ,  $s = 4$  and  $r = 1.6$ ), and suppose a social planner considers moving from  $\beta = 0$  to  $\beta = 1$  (i.e., from completely random to fully-meritocratic ranking). At  $\beta = 0$ , he/she considers the non-contribution equilibrium. At  $\beta = 1$ , he/she considers the high-contribution equilibrium. To assess which one he/she prefers, he/she makes a  $W_e$ -comparison. For this numerical example, it turns out that for any  $W_e$  with  $e < 10.3$  he/she prefers the high-contribution equilibria, while for a  $W_e$  with  $e \geq 10.3$ , he/she prefers the non-contribution equilibrium.<sup>12</sup> The following general statement about the welfare-optimal  $\beta$  can be made.

**Proposition 1.** For any  $R > \max\{mpcr, 1/(s-1)\}$ , there exists a population size  $n < \infty$  such that  $E[W_e(\phi); \beta] > E[W_e(\phi); \bar{\beta}]$  for all  $\beta \neq \bar{\beta}$  given any parameter of inequality aversion  $e$  bounded away from  $\infty$ .

**Proof.** Below  $\beta$ , the high-contribution equilibrium does not exist. The non-contribution equilibrium cannot be welfare-maximizing for any  $e < \infty$  as  $n \rightarrow \infty$ , because large efficiency gains (approaching infinity) would be foregone at virtually no inequality cost. Consider setting some  $\bar{\beta} \in (\beta, 1)$  instead, for which the high-contribution equilibrium exists. Write  $q_1^n$  for the probability of having more than one free-rider in any group for a realized outcome  $(\rho, \phi)$  given  $n < \infty$ . Since the number of free-riders does not increase as  $n$  increases,  $\partial q_1^n / \partial n < 0$ . Since contributors in groups with at most one free-rider receive a payoff strictly greater than one  $((s-1)R > 1)$ , we have  $E[W_e(\phi); \beta] > (1 - q_1^n) \times W_e(\phi_i = (s-1)R \forall i)$ . Because, given any  $\beta < 1$ ,  $\partial q_1^n / \partial n < 0$ , there therefore must exist a population size  $n < \infty$  above which  $E[W_e(\phi)] > W_e(\phi_i = 1 \forall i)$ . Hence,  $\beta$  is generally welfare-optimal for any  $e < \infty$  as  $n \rightarrow \infty$ .  $\square$

**Remark 2.**  $E[W_e(\phi); \beta] > E[W_e(\phi); \bar{\beta}]$  for all  $\beta \neq \bar{\beta}$  is also the case for  $n$  bounded away from infinity if (a)  $e$  is set below some bound  $e < \infty$  and/or (b)  $R$  is set above some bound  $R > 1/(s-1)$ .

<sup>12</sup> With  $e = 10.3$ ,  $W_e$  requires efficiency gains of more than twice the amount lost by any player to compensate for the additional inequality.



**Table 1.** Stem-and-leaf plot of individual payoffs for the non-contribution equilibrium (valid for any  $\beta \in [0, 1]$ ) and for the high-contribution equilibrium (when evaluated at  $\beta = 1$ ). Parameter values are  $n = 16$ ,  $s = 4$ ,  $r = 1.6$  and  $\beta = 1$ .

| High-Contribution Equilibrium<br>When $\beta = 1$ |                  | Payoff     | Non-Contribution Equilibrium<br>When $\beta = 0$ |  |
|---|------------------|------------|--|--|
| 13 14 ( $c_i = 1$ )                               | 0                | 0.0        | 0  |  |
|   | 0                | 0.2        | 0  |  |
|   | 0                | 0.4        | 0  |  |
|   | 0                | 0.6        | 0  |  |
|   | 2                | 0.8        | 0  |  |
|   | 0                | 1.0        | 16 ( $c_i = 0$ )                                 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 |
|   | 0                | 1.2        | 0  |  |
|   | 0                | 1.4        | 0  |  |
|   | 12 ( $c_i = 1$ ) | 1.6        | 0  |  |
|   | 2 ( $c_i = 0$ )  | 1.8        | 0  |  |
| 24.4  |                  | efficiency | 16   |  |

The stem of the table comprises the payoffs. The leafs are the number of players receiving that payoff (with their contribution decision) and the individual ranks of players corresponding to payoffs in the two equilibria. At the bottom, the efficiencies of the two outcomes are calculated.

#### 4.2. Heterogeneous Endowments

In the case of heterogeneous endowments and binary action space<sup>13</sup>, there is no equilibrium in which some contributors are grouped together with some defectors. Hence, no player is worse off in the high-contributing equilibria compared to the non-contributing one. Indeed, every contributive equilibria Pareto dominates the non-contributing one. The equilibrium where all but the  $S$  poorest players contribute is ex-post payoff dominant.

Hence, regardless of the value of the social planner preference parameter  $e$ , a social planner will always prefer the least possible amount of noise in the observations (ideally  $\beta = 1$ ) in order to ensure the most efficient outcome.

Note however that this does not mean that inequality does not increase due to the agents playing the efficient equilibrium. The starting inequality due to the initial endowments is amplified by the fact that the poorest players are the ones in the non-contributing group and thus the ones that do not benefit from the common good. Indeed, if the social planner were to measure inequality based on statistical dispersion measures rather than looking at the worst off player, he/she would observe an increase in inequality. As an example, consider the Gini coefficient<sup>14</sup>, a standard measure of inequality in a population, for a population of 16 players with initial endowments distributed according to a normal distribution  $\mathcal{N}(1, 0.1^2)$ : for a realization of the starting endowment resulting in a Gini coefficient of  $\sim 0.09$ , the Gini coefficient for the realized payoff in the ex-post dominant equilibrium is  $\sim 0.18$ . Hence, by this measure, the inequality in the population doubled after the game.

### 5. Logit Dynamics

In order to quantify the effect of noise, we simulate the above games with an agent-based model where players use logit-response [32,33] study the dynamics of the game. The following algorithm is used in these simulations:

<sup>13</sup> For continuous action space, the only equilibrium is non-contribution by all, and therefore, there is no social welfare analysis to be made.

<sup>14</sup> For a definition of the Gini coefficient, see, e.g., [30,31]

**Algorithm 1:** Logit choice dynamics.

- 
- 1 Initialization: Assign the initial endowments to each player sampling them from a Gaussian distribution with mean  $W_0$  and standard deviation  $\sigma$ . If some endowments are zero or negative, the sampling is repeated.<sup>a</sup>
  - 2 Initialization: Set initial strategy  $\alpha_i$  to 0  $\forall i$  (start the simulation in the fully-defective state<sup>b</sup>), and set time  $t$  to zero.
  - repeat**
  - 3   Update players' strategy:
    - for**  $i = 1$  **to**  $N$  **do**
    - Generate random number  $X$  uniformly in  $[0, 1)$ .
    - if**  $X < p$  **then**
    - Do not update player  $i$ 's strategy.<sup>c</sup>
    - else**
    - Player  $i$  chooses a new strategy  $j$  based on the logit probability distribution<sup>d</sup>:  

$$P_j = \frac{\exp(\lambda EU_j(\alpha_{-i}))}{\sum_k \exp(\lambda EU_k(\alpha_{-i}))}$$
, where  $\lambda$  is the rationality parameter and  $EU_j(\alpha_{-i})$  is the expected utility of strategy  $j$  given the  $\alpha_{-i}$ , the strategies played by all the other players in the previous round.
    - end**
    - end**
    - 4   Matching: Based on the contribution of each player, groups are formed as described in Section 2.1 and payoffs are then materialized.
    - 5   Reset wealth: Set the wealth of each player to his/her initial one and increase time  $t$  by one.
    - until**  $t < T$ ;

---

<sup>a</sup> The standard deviation is chosen for this to be an unlikely event.

<sup>b</sup> Different initial starting conditions have been explored, and they have been observed to have no effect on the final outcome of the simulation.

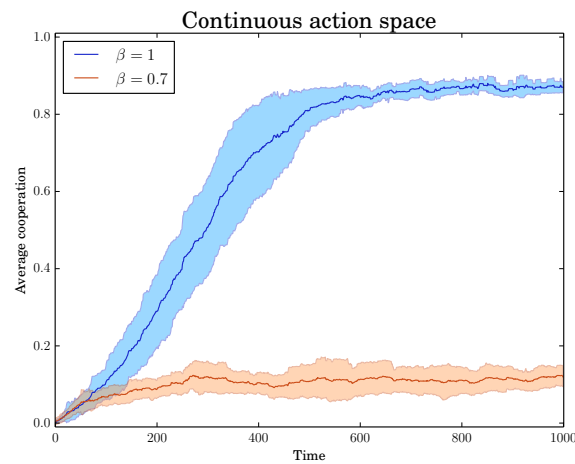
<sup>c</sup> Inertia was added to ensure the convergence to the pure-strategy Nash equilibrium (if existing). Without inertia, the best-response dynamics could oscillate around such an equilibrium.

<sup>d</sup> For a definition, see, e.g., [34].

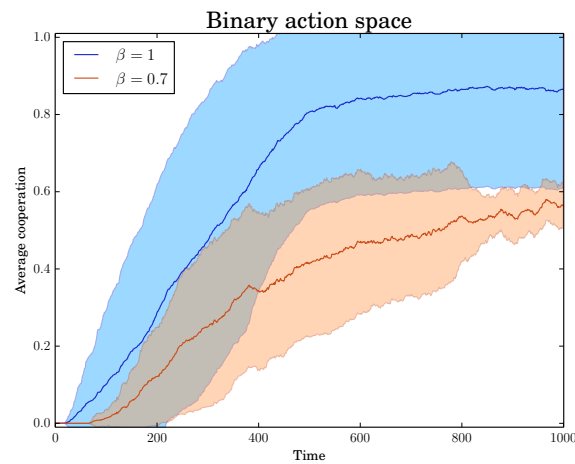
After  $T$  rounds, the algorithm stops, and the population average of the strategies is computed. The procedure is repeated  $EN$  times, and the ensemble average of the population average is obtained.

Figure 3 shows example results for homogeneous endowments in four different cases: perfect and imperfect observations for continuous and binary action spaces. When the matching mechanism is implemented perfectly, we observe a high level of contributions both in the binary and continuous action space (blue lines in Figure 3a,b). However, for an intermediate level of noise in the observed contributions, the results are quite different: in the binary case, we observe an intermediate level of average contributions (orange line in Figure 3b), while in the continuous case, we see very low amounts of cooperation (orange line in Figure 3a).

In Figure 4, we quantify how the percentage of contributions depends on the rate of return  $R$  and the level of noise in the matching mechanism  $\beta$  for agents with homogeneous endowments in a strategy space with continuous action. As predicted, the average level of cooperation weakly increases with increasing  $R$  and  $\beta$ , and we observe almost complete cooperation when approaching one. Naturally, regardless of the value of  $\beta$ , we observe no cooperation for a value of  $R \leq \frac{1}{s}$ . The contour lines indicate when the average contribution is above a certain threshold.

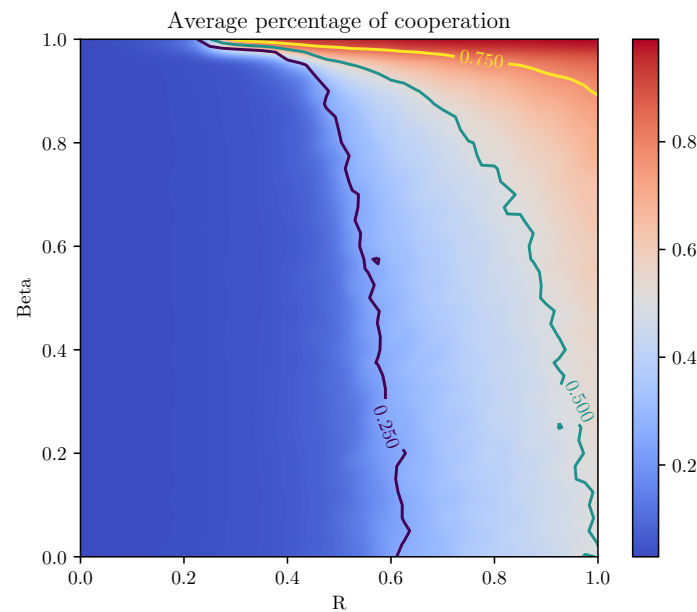


(a) Average cooperation in continuous action space. These simulations are obtained with  $N = 80$  and  $\lambda = 10$ .



(b) Average cooperation in binary action space. These simulations are obtained with  $N = 16$  and  $\lambda = 20$ .

**Figure 3.** The figure summarizes the average level of contributions in continuous and binary action spaces. The average is shown as a function of time for the voluntary contribution game with meritocratic matching with homogeneous endowments in four different cases: perfect and imperfect observations for continuous (a) and binary (b) action spaces. When the matching mechanism is implemented without noise ( $\beta = 1$ , blue lines in (a,b)), we observe high levels of cooperation in both cases. On the other hand, when the matching mechanism is implemented with partially-noisy observations  $\beta = 0.7$ , orange lines in (a,b), the result depends on the strategy space of the game: for a continuous action space (a), we observe a very low percentage of average contributions, while for a binary action space (b), we see a higher level of cooperation. The blurred areas around the lines represent the 95% confidence interval. The simulations were obtained for the following set of parameters:  $s = 4$ ,  $R = 0.5$ ,  $EN = 200$ ,  $p = 0.99$ . The simulations for the binary action space are obtained with a lower value for the rationality parameter in the logit function in order to speed up convergence and with a higher number of agents to reduce the error.



**Figure 4.** The figure depicts the average percentage of contributions as a function of the rate of return  $R$  and the level of noise in the matching mechanism  $\beta$ . For values of  $R \leq \frac{1}{s} = 0.25$  and/or too low values of  $\beta$ , we observe no cooperation. Otherwise, as predicted, the average level of cooperation weakly increases with increasing  $R$  and  $\beta$ , and we observe almost complete cooperation when approaching one. The contour lines indicate when the average contribution is above 25%, 50% and 75% of the whole population, respectively. The simulation was obtained for the following set of parameters:  $N = 80$ ,  $s = 4$ ,  $\lambda = 30$ ,  $EN = 100$ ,  $p = 0.99$ ,  $T = 2000$ .

## 6. Summary

Our analysis aimed to achieve three things. First, we study how the existence of high-contributions Nash equilibria depends on the strategy space of the players and on the homogeneity/heterogeneity of their initial endowment. Second, we allow noisy/fuzzy implementations of the mechanism everywhere between “no meritocracy” (completely random grouping) and “full meritocracy” (perfect contribution-based grouping). We found that the minimum meritocracy threshold that enables pure-strategy equilibria with high contributions decreases with the population size, the number of groups and with the rate of return. Where they exist, we established that, in the presence of some unconditional contributors, the high-contribution equilibria are selected rather than the non-contribution equilibrium. Mixed-strategy equilibria do not exist, unless the contribution space is binary (or very coarse).

Finally, we assessed the welfare properties of candidate equilibria to identify the welfare-maximizing regime, given varying degrees of inequality aversion of the regime. We found that setting meritocracy at the minimum meritocracy threshold that enables high-contribution equilibria maximizes welfare for general social welfare functionals as the population becomes infinitely large. An exception to this is in the case of coarse/binary action space and heterogeneous endowments where a mechanism with perfect matching is preferred regardless of how inequality averse the social planner is. For finite populations, the same result holds if (a) the inequality aversion is not extreme and (b) the rate of return is high.

More broadly, the mechanism we considered here belongs to a family of “grouping” mechanisms that enable high-contribution equilibria in voluntary contribution games.<sup>15</sup> Their common feature is that they implement non-random, contribution-based grouping, but require no payoff transfers between players. Importantly, these mechanisms incentivize contributions even if most of the players are narrowly self-interested. Our paper therefore complements other research on cooperative phenomena that arise from other-regarding preferences [43,44], in particular in public goods games (e.g., [45]).<sup>16</sup> We found that a fuzzy mechanism implementation could outperform a perfect implementation. It is an avenue for future research to consider fuzzy implementations of these alternative grouping mechanisms and of other mechanisms and to study other social dilemmas and games more generally under fuzzy mechanism design.

**Acknowledgments:** The authors would like to thank Ingela Alger, Lukas Bischofberger, Luis Cabral, Anna Gunnthorsdottir, Matthias Leiss, Michael Mäs, Bary Pradelski, Stefan Seifert, Peyton Young and Jörgen Weibull for helpful comments and discussions and the participants at Norms Actions Games 2014, the 25th and 26th International Conferences on Game Theory at Stony Brook 2015 and 2016, seminar participants at Bern 2016, Monte Verita 2015, PSE2015 and the Choice Group 2016, as well as an anonymous associate editor and two anonymous referees for helpful feedback. The authors also thank the Werner Oechslin Library Foundation. Nax and Duca acknowledge financial support by the European Commission through the ERC Advanced Investigator Grant ‘Momentum’ (Grant No. 324247). All remaining errors are ours.

**Author Contributions:** H.H.N. and S.D. did the modelling. All authors contributed substantially to all aspects of the work.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Pure Strategy Nash Equilibria

### Appendix A.1. Perfect Meritocracy

Let us consider the case of fully-heterogeneous agents playing the binary version of the voluntary contribution game.

In this case, we have  $c_i \in \{0, 1\}$ , and we define  $\mathbf{c} = \{c_i\}_{i \in N}$

We indicate with  $w_i$  the endowment of player  $i$ , and (without loss of generality) we order players such that  $w_1 > w_2 > \dots > w_n$ .

We will assume that the endowments of the players do not differ too much; in particular, we will assume that there exists a real finite number  $X$  such that  $w_i < w_{i-1} + X \forall i$ . If this assumption is not satisfied (or if the  $X$  is too big), then the only possible Nash equilibrium (NE) is non-contribution by all.

Formalism:

- We have a population of  $n$  players with  $N = \{1, 2, \dots, n\}$ .
- We have  $g$  groups,  $\{S_1, S_2, \dots, S_g\}$ , of size  $s$ .
- $\phi_i(c_i | \mathbf{c}_{-i})$  is the payoff of agent  $i$  with  $\mathbf{c}_{-i} \equiv \mathbf{c} \setminus \{c_i\}$ .

**Proposition A1.**  $\{c_i\} = 0 \forall i$  is an NE.

**Proof.** This is trivially proven: if every player but  $j$  were to contribute  $c_i = 0$ , the payoff for  $j$  would be the following:

$$\phi_j(c_j | \mathbf{c}_{-j}) = w_j(1 - c_j) + R w_j c_j$$

<sup>15</sup> Other mechanisms include: [35,36] consider endogenous group formation via voting; [37,38] study free group entry and exit; [39] analyze roommate-problem stable matching in pairwise-generated public goods; [40] study rematching based on reputation; [41,42] consider signaling.

<sup>16</sup> Using the terminology of [46], our paper therefore studies a ‘system’ rather than moral ‘acts’ or ‘intentions’. In our mechanism, the system assort contributions, i.e., actions, as other evolutionary biology mechanisms that lead to cooperation as, for example, kin selection (e.g., [47–49]), local interaction and/or assortative matching of preferences [50–54].

and for values of  $R$  less than one, the first order condition (FOC) results in  $c_j = 0$ .  $\square$

We will now study the conditions under which there can be “high efficient” Nash equilibria.

First of all, we note that a situation where the  $S$  wealthiest players contribute and all the others do not is a Nash equilibrium:

**Lemma A2.**  $\{c_i\} = 1 \forall i \in \{1 \dots S\}$  and  $\{c_i\} = 0 \forall i \notin \{1 \dots S\}$  is an NE if the rate of return  $R$  is greater than or equal to a threshold value  $\bar{R}$ .

**Proof.** Remark: It is enough to check that Player 1 (i.e., the wealthiest player) does not deviate from the contributive strategy because he/she is the player that benefits the least by contributing one.

The expected payoffs for Player 1 are:

$$E[\phi_1(c_1 = 1)] = R \left( \sum_{i=1}^S w_i \right)$$

and:

$$E[\phi_1(c_1 = 0)] = w_1 + \frac{R}{n-S+1} (w_2 + w_3 + w_4)$$

For  $\{c_i\} = 1 \forall i \in \{1 \dots S\}$  to be an NE, we need to have that  $E[\phi_1(c_1 = 1)] \geq E[\phi_1(c_1 = 0)]$ . Because of the decreasing wealth in the order of players, we have:

$$E[\phi_1(c_1 = 1)] > SRw_S \equiv A$$

and:

$$E[\phi_1(c_1 = 0)] < w_1 + \frac{(S-1)Rw_S}{n-S+1} \equiv B$$

Thus, it is enough to prove that  $A \geq B$ .

Since  $w_S > w_1 + (S-1)X$ , we can rewrite  $A \stackrel{?}{\geq} B$  as:

$$SR[w_1 - (S-1)X] - \frac{(S-1)}{n-S+1} [w_1 - (S-1)X] \stackrel{?}{\geq} w_1$$

which is equivalent to:

$$w_1(SR-1) - (S-1) \frac{Rw_1}{n-S+1} + SRX \left( 1 - \frac{2}{n-S+1} - S + \frac{S}{n-S+1} \right) + \frac{XR}{n-S+1} \geq 0$$

thus leading to:

$$R \geq \frac{w_1}{SR + SX \left( 1 - S + \frac{S-2}{n-S+1} \right) - \frac{(S-1)w_1 + X}{n-S+1}} \equiv \bar{R}$$

$\square$

Remark: Note that for  $X = 0$  (i.e., for homogeneous agents), the threshold  $\bar{R}$  reduces to the threshold found for homogeneous agents by [5].

We now proceed to show that all strategies (barring the one where everyone contributes) where the  $k \cdot S$  (with  $k$  an integer number) richest players contribute and the rest defect are Nash equilibria of the game:

**Lemma A3.** With  $m < n - S$  and  $k$  such that  $k \equiv m \pmod{S}$ :

If  $\{c_i\} = 1 \forall i \in \{1, \dots, k \cdot S\}$ ,  $c_j = 0$  else, is an NE, then  $\{c_i\} = 1 \forall i \in \{1, \dots, (k+1) \cdot S\}$ ,  $c_j = 0$  else, is also an NE.

Remark: Note that  $m < n - S$  excludes the situation where all players are contributing.

**Proof.** Because  $w_i > w_{i-1} > \dots$ , player  $kS$  can never be ranked lower than player  $kS$  if contributing. Hence, the remaining  $n - kS$  players are playing the same game as in Lemma A2, but with different numbers of players and groups:  $\tilde{n} = n - k \cdot S$  and  $\tilde{g} = g - k$ .

Hence, the same logic of Lemma A2 applies with the difference that, because  $n$  appears only in the numerator of  $\bar{R}(n)$ , the threshold  $\bar{\bar{R}}$  for the reduced game is smaller than the one for the full game; i.e.,  $\bar{\bar{R}} < \bar{R}$ .  $\square$

In all the equilibria described above, players in the last group left, i.e., the one with the  $S$  poorest player when all the other players contribute, will always defect:

**Lemma A4.**  $c_i = 0 \forall i \in \{k \cdot (S - 1) + 1, \dots, k \cdot S\}$ .

**Proof.** This follows trivially from the fact that with no possibility to be placed in a better group because of their low initial endowment, the poorest  $S$  players are playing the within the group best response where it is always true that  $E[\phi_i(c_i = 1)] < E[\phi_i(c_i = 0)]$ . Because  $R < 1$ , it is always the case that:

$$w_i + R \sum_{j \neq i} w_j > R \sum_j w_j$$

$\square$

The above lemmata lead to the conclusion that for binary action space and heterogeneous players, a maximum of  $g$  pure strategy Nash equilibria can coexist, depending on the value of the marginal per capita rate of return. Interestingly, for the equilibrium with  $k = 1$  to exist, the marginal rate of return has to be higher than for the equilibrium with  $k = 2$  to exist, and so on, until the lowest threshold for  $R$  such that the only contributive equilibrium is the one where all but one group is filled with contributors.

**Proposition A5.** *Depending on the value of the marginal rate of return  $R$ , there exists a maximum of  $g = \frac{S}{n}$  pure strategy Nash equilibria.*

*Their structure is so that the  $k \cdot s$  ( $0 \leq k < g$ ) poorest players defect and the  $(g - k) \cdot s$  richest players contribute.*

*The non-contribution equilibrium ( $k = 0$ ) always exists, regardless of the value of the marginal rate of return.*

*There can be no other pure strategy NE.*

**Proof.** Starting from the NE described in Lemma A2 and iteratively applying Lemma A3, we can derive the existence of the  $g - 1$  contributive equilibria. Note that the threshold for the marginal rate of return derived in Lemma A2 is the highest of all the thresholds.

Proposition A1 shows that the non-contribution equilibrium always exists regardless of the value of  $R$ .

The heterogeneity of players' endowment ensures that there can be no mixed group such that contributors and non-contributors are grouped together. This, together with Lemma A4, proves that there can be no other pure strategy Nash equilibria other than the ones described above.  $\square$

Let us now move to continuous action space and show that, in the case of homogeneous agents, when some strategies in equilibrium have positive contributions, there can be no player  $i$  contributing  $0 < c_i < 1$ .

**Lemma A6.** *If in equilibrium there is some player  $i$  for which  $c_i$  s.t.  $c_i > 0$ , then there cannot be any player  $j$  contributing  $0 < c_j < 1$ .*



**Proof.** If there are no players other than player  $j$  contributing, then from Proposition A7, it follows that  $j$ 's best reply is to defect, as well.

If there are other players contributing more than zero, let us call  $j$  the player(s) with the highest contribution.

If there are no ties regarding group membership,  $j$  could reduce his/her contribution to  $c_j - \varepsilon$ , with  $\varepsilon$  arbitrarily small, and remain in the same group. If there are ties,  $j$  could increase his/her contribution by an arbitrary small amount and be sure to be grouped together with players contributing. Similarly to Lemma 3 in [5], it is possible to prove that for an arbitrary small  $\varepsilon$ , we have:

$$E[\phi(c_j)] < E[\phi(c_j + \varepsilon)]$$

and hence that  $c_j \neq \{0, 1\}$  cannot be an equilibrium strategy.  $\square$

## Appendix A.2. Fuzzy Mechanism

We now turn our attention to Nash equilibria in the case of the implementation of the fuzzy mechanism.

**Proposition A7.** For any population size  $n > s$ , group size  $s > 1$ , rate of return  $r \in (1, s)$  and meritocratic matching factor  $\beta \in [0, 1]$ , universal non-contribution is always a Nash equilibrium.

The proof of Proposition A7 follows from the fact that, given any  $\beta$  and for  $c_{-i}$  such that  $\sum_{j \neq i} c_j = 0$ , we have:

$$1 = E[\phi_i(0|c_{-i})] > E[\phi_i(1|c_{-i})] = R. \quad (A1)$$

Equation (A1), in other words, means that it is never the best response to be the only contributor for any level of  $\beta$ . If, for any level of  $\beta$ , given any  $c_{-i}$ ,  $E[\phi_i(0|c_{-i})] > E[\phi_i(1|c_{-i})]$  holds for all  $i$ , then we have a situation where non-contribution is the strictly dominant strategy. In that case, for any level of meritocracy ( $\beta$ ), universal non-contribution is the unique Nash equilibrium.

Lemma A6 implies that for players with homogeneous endowments, it is enough to focus on the extremal strategies when analyzing Nash equilibria.

We shall now proceed to show that additional high-contribution equilibria exist if the marginal per capita rate of return ( $R$ ) and the meritocratic matching fidelity ( $\beta$ ) are high enough. Before we do this, we write  $1^m$  to refer to a strategy profile where “ $m$  players contribute, all others free-ride” and  $1_{-i}^m$  for the same statement excluding player  $i$ , i.e., “ $m$  of the other players contribute,  $n - m$  other players free-ride”. We also define a critical threshold for the marginal per capita rate of return of  $m_{pcr} = \frac{n-s+1}{ns-s^2+1}$ , which is the threshold, as identified by [5], for which the high-contribution equilibria exist for the case of perfect merit (when  $\beta = 1$ ) and homogeneous endowments.

**Proposition A8.** Given population size  $n > s$ , group size  $s > 1$  and rate of return  $r$  such that  $R \in (m_{pcr}, 1)$ , there exists a necessary meritocracy level,  $\beta \in (0, 1)$ , above which there is a high-contribution Nash equilibrium, where  $m > 0$  agents contribute and the remaining  $n - m$  agents free-ride.

**Proof.** The following two conditions must hold for Proposition A8 to be true:

$$E[\phi_i(1|1_{-i}^m)] \geq E[\phi_i(0|1_{-i}^{m-1})] \quad (A2)$$

$$E[\phi_i(0|1_{-i}^m)] \geq E[\phi_i(1|1_{-i}^{m+1})] \quad (A3)$$

The proof for the existence of an equilibrium in which some appropriate (positive) number of contributors  $m$  exists for the case when  $\beta = 1$  and  $R \geq m_{pcr}$  follows from Theorem 1 in [5], in which case both Equations (A2) and (A3) are strictly satisfied.

The fixed point argument behind that result becomes clear by inspection of Terms (ii) and (iii) in Expression (2): namely, the decision to contribute rather than to free-ride is a trade-off between (ii) ‘the sure loss on own contribution’, which is zero for free-riding, versus (iii) ‘the expected return on others’ contributions’, which may be larger by contributing rather than by free-riding depending on how many others also contribute. Obviously, when  $c_{-i}$  is such that  $\sum_{j \neq i} c_j = 0$  or  $\sum_{j \neq i} c_j = (n - 1)$  (i.e., if either all others free-ride or all others contribute), it is the case that  $\phi_i(0|c_{-i}) > \phi_i(1|c_{-i})$ . Hence, in equilibrium,  $0 < m < n$ .

Now, suppose  $1^m$  describes a pure-strategy Nash equilibrium for  $\beta = 1$  with  $0 < m < n$  and  $R \in (mpcr, 1)$  in which case Equations (A2) and (A3) are strictly satisfied. Note that  $\beta$  has a positive effect on the expected payoff of contributing and a negative effect on the expected payoff of free-riding:

$$\partial E [\phi_i(1|1_{-i}^m)] / \partial \beta > 0 \quad (\text{A4})$$

$$\partial E [\phi_i(0|1_{-i}^m)] / \partial \beta < 0 \quad (\text{A5})$$

When  $\beta = 0$ , we know that  $\phi_i(1|1_{-i}^m) = R < \phi_i(0|1_{-i}^m) = 1$  for any  $m$ . However, by existence of the equilibrium with  $m > 0$  contributors when  $\beta = 1$ , provided that  $R > mpcr$  is satisfied, there must exist some maximum value of  $\beta \in (0, 1)$ , at which either Equation (A2) or Equation (A3) first binds due to continuity of Expressions (A4) and (A5) in  $\beta$ . That level is the bound on  $\beta$  above which the pure-strategy Nash equilibrium with  $m > 0$  exists.  $\square$

**Remark A9.** Note that, for a finite population of size  $n$ , a group size  $s$  larger than one implies that  $mpcr > 1/s$  for Proposition A8 to be true, but as  $n \rightarrow \infty$ ,  $mpcr$  converges to  $1/s$ .<sup>17</sup>

A special case of such a pure-strategy Nash equilibrium is the high-contribution Nash equilibrium as identified by (see [5]): in our setup, the almost-no-free-riders pure-strategy Nash equilibrium generalizes to the pure-strategy Nash equilibrium in which  $m$  is chosen to be the largest value given  $n, s, r$  for which Equations (A2) and (A3) hold. For that  $m$  to be larger than zero,  $\beta$  needs to be larger than  $\beta$  (Proposition (A8)).

## Appendix B. Mixed-Strategy Nash Equilibria

Now, we shall compare the asymmetric high-contribution Nash equilibria with potential symmetric mixed-strategy Nash equilibria. For this, we define  $p_i \in [0, 1]$  as a mixed strategy with which player  $i$  plays ‘contributing’ ( $c_i = 1$ ), while playing ‘free-riding’ ( $c_i = 0$ ) with  $(1 - p_i)$ . Write  $p = \{p_i\}_{i \in N}$  for a vector of mixed strategies. Write  $1^p$  for “all players play  $p$ ” and  $1_{-i}^p$  for the same statement excluding some player  $i$ .

First, we consider the case of a binary contribution space, i.e., when contributions can only be full or null.

**Proposition A10.** Consider the case of  $c_i \in \{0, 1\}$  for all  $i$ . Given population size  $n > s$  and group size  $s > 1$ , there exists a rate of return  $r$  such that  $R \in [mpcr, 1)$  beyond which there exists a necessary meritocracy level,  $\beta \in (0, 1)$ , such that there always are two mixed strategy profiles, where every agent places weight  $p > 0$  on contributing and  $1 - p$  on free-riding, that constitute a symmetric mixed-strategy Nash equilibrium. One will have a high  $\bar{p}$  (the near-efficient symmetric mixed-strategy Nash equilibrium), and one will have a low  $p$  (the less-efficient symmetric mixed-strategy Nash equilibrium).

<sup>17</sup> It is easy to check that  $\lim_{n \rightarrow \infty} mpcr = 1/s$ .

**Proof.** The symmetric mixed-strategy Nash equilibrium exists if there exists a  $p \in (0, 1)$  such that, for any  $i$ ,

$$\mathbf{E} [\phi_i(0|1_{-i}^p)] = \mathbf{E} [\phi_i(1|1_{-i}^p)], \quad (\text{A6})$$

because, in that case, player  $i$  has the best response also playing  $p_i = p$ , guaranteeing that  $1^p$  is a Nash equilibrium. Proposition A8 implies that, if  $R > mpcr$ , Equations (A2) and (A3) are strictly satisfied when  $\beta = 1$  for  $m$  contributors corresponding to the almost-no-free-riders pure-strategy Nash equilibrium. Indeed, Expressions (A2) and (A3) imply lower and upper bounds (see [5]) on the number of free-riders given by:

$$l = \frac{n - nR}{1 - R + nR - r}, \quad u = 1 + \frac{n - nR}{1 - R + nR - r}. \quad (\text{A7})$$

Part 1. First, we will show, given any game with population size  $n$  and group size  $s$ , for the case when  $\beta = 1$ , that there is (i) at least one symmetric mixed-strategy Nash equilibrium when  $R \rightarrow 1$ ; (ii) possibly none when  $R = mpcr$ ; and (iii) a continuity in  $R$  such that there is some intermediate value of  $R \in [mpcr, 1)$  above which at least one symmetric mixed-strategy Nash equilibrium exists, but not below.

(i) Because  $\partial \mathbf{E} [\phi_i(c_i|1_{-i}^p)] / \partial p > 0$  for all  $c_i$ , there exists a  $p \in (\frac{m-1}{n}, \frac{m+1}{n})$  such that Expression (A6) holds if  $R \rightarrow 1$ . This is the standard symmetric mixed-strategy Nash equilibrium, which always exists in a symmetric two-action  $n$ -person game where the only pure-strategy equilibria are asymmetric and of the same kind as the high-contribution pure-strategy Nash equilibrium (see the proof of Theorem 1 in [55]). In this case, the presence of the non-contribution Nash equilibrium makes no difference because the incentive to free-ride vanishes as  $R \rightarrow 1$ .

(ii) If  $R = mpcr$ , one or both of the equations, (A2) or (A3), bind. Hence, unless Expression (A6) holds exactly at  $p = m/n$  (which is a limiting case in  $n$  that we will address in Proposition A15), there may not exist any  $p$  such that Expression (A6) holds. This is because the binomially-distributed proportions of contributors implied by  $p$ , relatively speaking, place more weight on the incentive to free-ride than to contribute because universal free-riding is consistent with the non-contribution Nash equilibrium, while universal contributing is not a Nash equilibrium. In this case, the incentive to free-ride is too large for a symmetric mixed-strategy Nash equilibrium to exist.

(iii)  $\partial \mathbf{E} [\phi_i(c_i|1_{-i}^p)] / \partial r$  is a different linear, positive constant for both  $c_i = 0$  and  $c_i = 1$ . At and above some intermediate value of  $R$ , therefore, there exists a  $p \in (0, 1)$  such that, if played in a symmetric mixed-strategy Nash equilibrium, the incentive to free-ride is mitigated sufficiently to establish Equation (A6). We shall refer to this implicit minimum value of  $R$  by  $\overline{mpcr}$ .

Finally, for any  $p > 0$  constituting a symmetric mixed-strategy Nash equilibrium when  $\beta = 1$ ,  $\mathbf{E} [\phi_i(0|1_{-i}^p)] = \mathbf{E} [\phi_i(1|1_{-i}^p)] > 1$ . Because of this, a similar argument as in Proposition A8 applies to ensure the existence of some  $\beta \in (0, 1)$  above which the symmetric mixed-strategy Nash equilibrium continues to exist when  $R > \overline{mpcr}$ : because, at  $\beta = 1$ , Equations (A2) and (A3) are strictly satisfied and  $\mathbf{E} [\phi_i(0|1_{-i}^p)] = \mathbf{E} [\phi_i(1|1_{-i}^p)] > 1$ , there therefore must exist some  $\beta < 1$  and  $p' < p$  satisfying Equation (A6) while still satisfying  $\mathbf{E} [\phi_i(0|1_{-i}^{p'})] = \mathbf{E} [\phi_i(1|1_{-i}^{p'})] > 1$ . Note that this implicit bound here may be different from that in Proposition A8.

Part 2. If  $R > \overline{mpcr}$  and  $\beta > \beta$ , the existence of two equilibria with  $\bar{p} > p > 0$  is shown by analysis of the comparative statics of Equation (A6).

First note that, for any  $R > \overline{mpcr}$  and  $\beta > \beta$ ,  $\partial \mathbf{E} [\phi_i(0|1_{-i}^p)] / \partial \beta < 0$ , while  $\partial \mathbf{E} [\phi_i(1|1_{-i}^p)] / \partial \beta > 0$ .  $p$  therefore has to take different values for Equation (A6) to hold for two different values of  $\beta$  above  $\beta$ . It is unclear whether it has to take a higher or lower value. Note also that both  $\partial \mathbf{E} [\phi_i(0|1_{-i}^p)] / \partial p > 0$

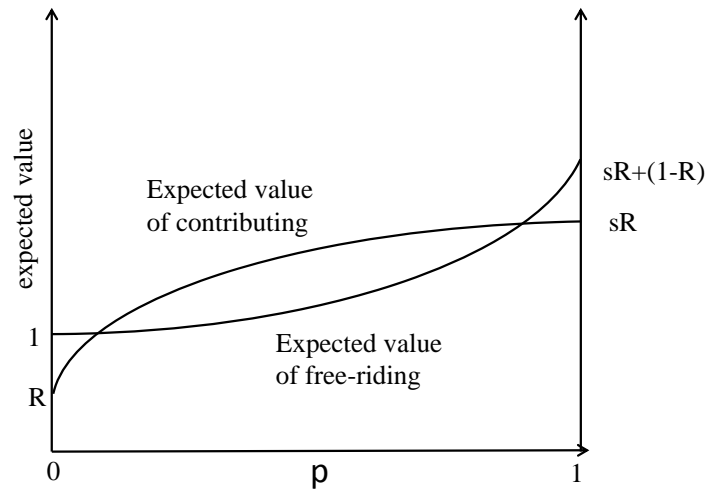
and  $\partial \mathbf{E} [\phi_i(1|1_{-i}^p)] / \partial p > 0$  for all  $\beta \in (0, 1)$ . We can rearrange the partial derivative with respect to  $\beta$  of Expression A6, and obtain:

$$\partial p / \partial \beta = \frac{\partial \mathbf{E} [\phi_i(1|1_{-i}^p)] / \partial \beta - \partial \mathbf{E} [\phi_i(0|1_{-i}^p)] / \partial \beta}{\partial \mathbf{E} [\phi_i(0|1_{-i}^p)] / \partial p - \partial \mathbf{E} [\phi_i(1|1_{-i}^p)] / \partial p}. \quad (\text{A8})$$

Expression A8 is negative if the denominator is negative, because the numerator is always positive.

**Claim A11.** *The denominator of Equation (A8) is negative when  $p$  is low, and positive when  $p$  is high.*

Write  $\bar{w}_{c_i}^i$  and  $w_{c_i}^i$  respectively for the probabilities with which agent  $i$  is matched in an above- or below-average group when playing  $c_i$  where the average is taken over contributions excluding  $i$ . Write  $\mathbf{E} [\bar{\phi}_i(c_i|1_{-i}^p)]$  and  $\mathbf{E} [\phi_i(c_i|1_{-i}^p)]$  for the corresponding expected payoffs.



**Figure A1.** Expected payoffs of contributing versus free-riding. The expected values of  $\phi_i(0|1_{-i}^p)$  and  $\phi_i(1|1_{-i}^p)$  are plotted as functions of the probability  $p$  for some fixed  $\beta > \beta$ . The two planes intersect at the bifurcating symmetric mixed-strategy Nash equilibrium-values of  $\bar{p}$  and  $p$  (see Proposition A10). The relative slopes of the two curves illustrate the proposition. Note that this figure is a slice through Figure A2 along a value of  $\beta > \beta$ .

Recall that, for  $\beta > 0$  and  $1_{-i}^p \in (0, 1)$ , Expression (A12)<sup>18</sup> holds, where  $\hat{k}$  is compatible with a perfect ordering  $\hat{\pi}$ , and  $\tilde{k}$  is any rank compatible with a mixed ordering  $\tilde{\pi}$ . When  $1_{-i}^p = 0$  or  $1_{-i}^p = 1$ , the probability of agent  $i$  to take rank  $j$ ,  $f_{ij}^\beta$ , depends on his/her choice of  $c_i$ , but  $\bar{w}_{c_i}^i = w_{c_i}^i = 0$  for any choice of contribution  $c_i$ .

For  $p(0, 1)$ , we shall rewrite  $\partial \mathbf{E} [\phi_i(0|1_{-i}^p)] / \partial p$  in the denominator of Equation (A8) as:

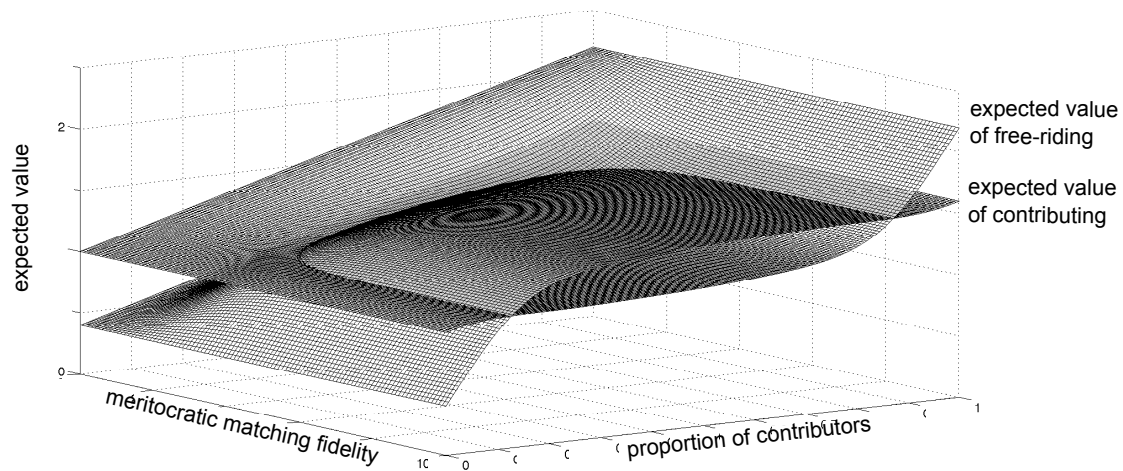
$$\frac{\partial}{\partial p} [\bar{w}_0^i * \mathbf{E} [\bar{\phi}_i(0|1_{-i}^p)]] + \frac{\partial}{\partial p} [w_0^i * \mathbf{E} [\phi_i(0|1_{-i}^p)]] \quad (\text{A9})$$

<sup>18</sup> See Appendix C.

and  $\partial \mathbf{E} [\phi_i(1|1_{-i}^p)] / \partial p$  as:

$$\frac{\partial}{\partial p} [\bar{w}_1^i * \mathbf{E} [\bar{\phi}_i(1|1_{-i}^p)]] + \frac{\partial}{\partial p} [w_1^i * \mathbf{E} [\phi_i(1|1_{-i}^p)]] . \quad (\text{A10})$$

Notice that, for large  $\beta$ ,  $w_0^i \gg \bar{w}_0^i$  when  $p$  is close to zero, and  $w_1^i \ll \bar{w}_1^i$  when  $p$  is close to one. Moreover, notice that the existence of the pure-strategy Nash equilibrium with high contribution for high levels of  $\beta$  ensures that  $\mathbf{E} [\phi_i(0|1_{-i}^p)]$  is not always larger than  $\mathbf{E} [\phi_i(1|1_{-i}^p)]$ . It therefore follows from continuity in  $\beta$  that Expression A10 exceeds Expression A9 when  $p$  is low and that Expression A9 exceeds Expression A10 when  $p$  is high; hence, the denominator of Equation A8 is negative when  $p$  is low and positive when  $p$  is high. Figure A1 illustrates.  $\square$



**Figure A2.** The figure depicts the expected payoffs of contributing versus free-riding for the economy with  $n = 16$ ,  $s = 4$ ,  $r = 1.6$ . The expected values of  $\phi_i(0|1_{-i}^p)$  and  $\phi_i(1|1_{-i}^p)$  are plotted as functions of contribution probability  $p$  and meritocratic matching fidelity  $\beta$ . The two planes intersect at the bifurcating symmetric mixed-strategy Nash equilibrium-values of  $\bar{p}$  and  $p$  (see Proposition A10). Notice that the expected values of both actions increase linearly in  $p$  when the meritocratic matching fidelity is zero, but turn increasingly S-shaped for larger values, until they intersect at  $\bar{p}$  and  $p$ .

**Remark A12.** Note that the necessary meritocracy level  $\beta$  in Propositions A8 and A10 need not be the same. We shall write  $\beta$  for whichever level is larger.

**Corollary A13.** For intermediate values of  $p$ , contributing is a better reply than free-riding.

**Corollary A14.** For the case when  $c_i \in [0, 1]$ , a symmetric mixed-strategy Nash equilibrium does not exist.

**Proof.** Consider the symmetric mixed-strategy Nash equilibrium from the above proof with  $p_i = p$  for all  $i$ . Suppose all  $j \neq i$  play  $p$ . Then, in the neighborhood of  $\epsilon = 0$ ,  $\partial \mathbf{E} [\phi_i(\epsilon|1_{-i}^p)] / \partial \epsilon > 0$ , because playing  $\epsilon > 0$  for this player will always rank him above others playing zero. Hence,  $p_i = p$  for all  $i$  cannot be an equilibrium.  $\square$

**Proposition A15.** Given group size  $s > 1$ , then, if  $\beta = 1$ , as  $n \rightarrow \infty$ : (i)  $1^m/n$  of the high-contribution pure-strategy Nash equilibrium and  $\bar{p}$  of the near-efficient symmetric mixed-strategy Nash equilibrium converge; and (ii) the range of  $R$  for which these equilibria exist converges to  $(1/s, 1)$ .

**Proof.** Suppose  $R > \overline{m p c r}$ , i.e., that both symmetric mixed-strategy Nash equilibrium and high-contribution pure-strategy Nash equilibrium exist. Let  $1^m$  describe the high-contribution pure-strategy Nash equilibrium and  $1^p$  describe the near-efficient symmetric mixed-strategy Nash equilibrium. Recall that expressions under (A7) summarize the lower and upper bound on the number of free-riders,  $(n - m)$  in the high-contribution pure-strategy Nash equilibrium. Taking  $\lim_{n \rightarrow \infty}$  for those bounds implies a limit lower bound of  $\frac{1}{1+n \frac{R-r/n}{1-R}}$  and a limit upper bound of the expected proportion of free-riders of  $\frac{1}{n} + \frac{1}{1+n \frac{R-r/n}{1-R}}$  and, thus, bounds on the number of free-riders that contain at most two integers and at least one free-rider (notice that the limits imply that exactly one person free-rides as  $R \rightarrow 1$ ). We know that, if there is one more free-rider than given by the upper bound, then Equation (A3) is violated. Similarly, if there is one fewer free-rider than given by the lower bound, then Equation (A2) is violated.

With respect to the near-efficient symmetric mixed-strategy Nash equilibrium, recall that Expression A6 must hold; i.e.,  $E[\phi_i(0|1^p_{-i})] = E[\phi_i(1|1^p_{-i})]$ . We can rewrite  $E[\phi_i(c_i|1^p_{-i})]$  as  $E[\phi_i(c_i|B)]$ , where  $B$  is the number of other players actually contributing (playing  $c_i = 1$ ), which is distributed according to a binomial distribution  $Bin(p, n)$  with mean  $E[B] = np$  and variance  $V[B] = np(1-p)$ . As  $n \rightarrow \infty$ , by the law of large numbers, we can use the same bounds obtained for the high-contribution pure-strategy Nash equilibrium to bound  $(B/n) \in [(n-u)/n, (n-l)/n]$ , which converges to the unique  $p$  for which Expression (A6) actually holds.<sup>19</sup>

Suppose all players contribute with probability  $p$  corresponding to the near-efficient symmetric mixed-strategy Nash equilibrium limit value. Then,  $\lim_{n \rightarrow \infty} V[(B/n)] = \lim_{n \rightarrow \infty} \frac{p(1-p)}{n} = 0$  for the actual proportion of contributors. Hence, the limit for the range over  $R$  necessary to ensure existence converges to that of the high-contribution pure-strategy Nash equilibrium, which by Remark A9 is  $(1/s, 1)$ .  $\square$

**Remark A16.** In light of the limit behavior, it is easy to verify, *ceteris paribus*, that the value of the marginal per capita rate of return necessary to ensure the existence of the symmetric mixed-strategy Nash equilibrium is decreasing in population size  $n$ , but increasing in group size  $s$ ; i.e., decreasing in relative group size  $s/n$ .

## Appendix C. Properties of the Fuzzy Ranking

Fixing  $\beta \in [0, 1]$ , we indicate with  $\bar{k}_i^\beta(c_i)$  the expected ranking of player  $i$  when contributing  $c_i$ . The ranking is such that, under a perfect ordering,  $k_i > k_j \Rightarrow c_i \leq c_j$ . We define the fuzziness such that for each player, the observed contribution  $x_i$  is distributed normally around the true contribution  $c_i$  with standard deviation  $\sigma$ . We define  $\sigma^2 = \frac{1-\beta}{\beta}$  such that  $x_i \sim \mathcal{N}(c_i, \sigma^2)$ .

The following properties hold:

- For  $\beta = 1$ , the contributions are perfectly observable, and the ranking follows a perfect ordering.
- For  $\beta = 0$ , contributions do not matter, and the ordering of the players is determined completely at random. This follows from the property of the normal distribution for  $\sigma \rightarrow \infty$  ( $\beta \rightarrow 0$ ).
- For  $0 < \beta < 1$ , the expected ranking for every player has the following properties:

$$\frac{\partial E[\bar{k}_i^\beta(c_i)]}{\partial c_i} < 0 \quad (\text{A11})$$

$$E[\bar{k}_i^\beta(c_L) - \bar{k}_i^\beta(c_H)] > 0 \quad ; \quad c_L < c_H \quad (\text{A12})$$

<sup>19</sup> Details concerning the use of the law of large numbers can be followed based on the proof in [55].



$$\frac{\partial \mathbf{E} [\bar{k}_i^\beta (c_L) - \bar{k}_i^\beta (c_H)]}{\partial \beta} > 0 \quad (\text{A13})$$

Proofs

Let us define  $x_L$  and  $x_H$  as the imperfectly observed contributions when a player contributes  $c_L$  and  $c_H$ , respectively; i.e.,  $x_L \sim \mathcal{N}(c_L, \sigma^2)$  and  $x_H \sim \mathcal{N}(c_H, \sigma^2)$ .

**Lemma A17.** *Proof of (A11):*

**Proof.** This follows trivially from  $\mathbf{E} [\bar{k}_i^\beta (c_q)] = \bar{k}_i^\beta (\mathbf{E} [c_q])$  and the matching rules.  $\square$

**Lemma A18.** *Proof of (A12):*

**Proof.** It again follows from  $\mathbf{E} [\bar{k}_i^\beta (c_q)] = \bar{k}_i^\beta (\mathbf{E} [c_q])$ , and hence:

$$\mathbf{E} [\bar{k}_i^\beta (c_L) - \bar{k}_i^\beta (c_H)] = \mathbf{E} [\bar{k}_i^\beta (c_L)] - \mathbf{E} [\bar{k}_i^\beta (c_H)] = \bar{k}_i^1 (c_L) - \bar{k}_i^1 (c_H) > 0 \text{ for } c_L < c_H. \quad \square$$

**Lemma A19.** *Proof of (A13):*

**Proof.** Let us write Equation (A13) as a function of  $\sigma$ :

$$\frac{\partial g}{\partial \beta} = -\frac{1}{\beta^2} \frac{\partial g}{\partial \sigma^2} = -\frac{1}{2\sigma\beta^2} \frac{\partial g}{\partial \sigma}$$

with a slight abuse of notation. Hence, we have that:

$$\frac{\partial \mathbf{E} [\bar{k}_i^\beta (c_L) - \bar{k}_i^\beta (c_H)]}{\partial \beta} > 0 \Rightarrow \frac{\partial \mathbf{E} [\bar{k}_i^\beta (c_L) - \bar{k}_i^\beta (c_H)]}{\partial \sigma} < 0.$$

The expected difference in ranking between  $c_L$  and  $c_H$  can be rewritten in terms of the probability of observing a higher realized contribution for  $c_L$  than for  $c_H$ ; i.e.,

$$\frac{\partial \mathbf{E} [\bar{k}_i^\beta (c_L) - \bar{k}_i^\beta (c_H)]}{\partial \sigma} < 0 \Leftrightarrow \frac{\partial P(z > 0)}{\partial \sigma} < 0$$

with  $z \equiv x_L - x_H$ . Note that  $z \sim \mathcal{N}(c_L - c_H, 2\sigma^2)$ <sup>20</sup>.

By definition, we have:

$$P(z > 0) = 1 - P(z \leq 0) \equiv 1 - \Phi\left(\frac{c_H - c_L}{\sqrt{2}\sigma}\right) = \frac{1}{2} \left[1 - \operatorname{erf}\left(\frac{c_H - c_L}{2\sigma}\right)\right]$$

where  $\Phi(x)$  denotes the cumulative distribution function of the normal distribution and  $\operatorname{erf}(x)$  the error function.

We thus obtain:

$$\frac{\partial \mathbf{E} [\bar{k}_i^\beta (c_L) - \bar{k}_i^\beta (c_H)]}{\partial \beta} > 0 \Leftrightarrow \frac{\partial P(z > 0)}{\partial \sigma} < 0 \Leftrightarrow \frac{\partial \operatorname{erf}\left(\frac{c_H - c_L}{2\sigma}\right)}{\partial \sigma} > 0$$

<sup>20</sup> See, e.g., [56]



However,  $\operatorname{erf}(x) = \int_0^x e^{-t^2} dt$ , and hence,  $\frac{\partial \operatorname{erf}(x)}{\partial x} = e^{-x^2}$ . We thus obtain that:

$$\frac{\partial \operatorname{erf}\left(\frac{c_H - c_L}{2\sigma}\right)}{\partial \sigma} = e^{-\left(\frac{c_H - c_L}{2\sigma}\right)^2} \cdot 2 \left(\frac{c_H - c_L}{2\sigma}\right) \cdot \frac{c_H - c_L}{2\sigma^2} > 0$$

which is always positive for  $\beta > 0$ .  $\square$

## References

1. Abreu, D.; Pearce, D.; Stacchetti, E. Toward a theory of discounted repeated games with imperfect monitoring. *Econom. J. Econom. Soc.* **1990**, *58*, 1041–1063.
2. Fudenberg, D.; Levine, D.; Maskin, E. The folk theorem with imperfect public information. *Econom. J. Econom. Soc.* **1994**, *62*, 997–1039.
3. Wu, J.; Axelrod, R. How to cope with noise in the iterated prisoner's dilemma. *J. Confl. Resolut.* **1995**, *39*, 183–189.
4. Alchian, A.A.; Demsetz, H. Production, information costs, and economic organization. *Am. Econ. Rev.* **1972**, *62*, 777–795.
5. Gunnthorsdottir, A.; Vragov, R.; Seifert, S.; McCabe, K. Near-efficient equilibria in contribution-based competitive grouping. *J. Public Econ.* **2010**, *94*, 987–994.
6. Andreoni, J. Why free ride? Strategies and learning in public goods experiments. *J. Public Econ.* **1988**, *37*, 291–304.
7. Andreoni, J. An experimental test of the public goods crowding-out hypothesis. *Am. Econ. Rev.* **1993**, *83*, 1317–1327.
8. Andreoni, J. Cooperation in public-goods experiments: kindness or confusion? *Am. Econ. Rev.* **1995**, *85*, 891–904.
9. Palfrey, T.R.; Prisbrey, J.E. Altruism, reputation and noise in linear public goods experiments. *J. Public Econ.* **1996**, *61*, 409–427.
10. Palfrey, T.R.; Prisbrey, J.E. Anomalous behavior in public goods experiments: How much and why? *Am. Econ. Rev.* **1997**, *87*, 829–846.
11. Goeree, J.K.; Holt, C.A.; Laury, S.K. Private costs and public benefits: Unraveling the effects of altruism and noisy behavior. *J. Public Econ.* **2002**, *83*, 255–276.
12. Ferraro, P.J.; Vossler, C.A. The source and significance of confusion in public goods experiments. *BE J. Econ. Anal. Policy* **2010**, *10*, 53.
13. Fischbacher, U.; Gaechter, S. Social preferences, beliefs, and the dynamics of free riding in public good experiments. *Am. Econ. Rev.* **2010**, *100*, 541–556.
14. Bayer, R.C.; Renner, E.; Sausgruber, R. Confusion and learning in the voluntary contributions game. *Exp. Econ.* **2013**, *16*, 478–496.
15. Nax, H.H.; Burton-Chellew, M.N.; West, S.A.; Young, H.P. Learning in a Black Box. *J. Econ. Behav. Organ.* **2016**, *127*, 1–15.
16. Burton-Chellew, M.N.; Nax, H.H.; West, S.A. Payoff-based learning explains the decline in cooperation in public goods games. *Proc. R. Soc. Lond. B Biol. Sci.* **2015**, *282*, 20142678.
17. Chaudhuri, A. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Exp. Econ.* **2011**, *14*, 47–83.
18. Ledyard, J.O. Public Goods: A Survey of Experimental Research. In *Handbook of Experimental Economics*; Kagel, J.H., Roth, A.E., Eds.; California Institute of Technology: Pasadena, CA, USA, 1995; Volume 37, pp. 111–194.
19. Gunnthorsdottir, A.; Thorsteinsson, P. Tacit Coordination and Equilibrium Selection in a Merit-Based Grouping Mechanism: A Cross-Cultural Validation Study. In Proceedings of the 19th International Congress on Modelling and Simulation (MODSIM2011), Perth, Australia, 12–16 December 2011.
20. Rabanal, J.P.; Rabanal, O.A. *Efficient Investment via Assortative Matching: A Laboratory Experiment*; Hamilton College: Clinton, NY, USA. Unpublished work, 2014.
21. Ones, U.; Putterman, L. The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *J. Econ. Behav. Organ.* **2007**, *62*, 495–521.
22. Duca, S.; Helbing, D.; Nax, H.H. Assortative Matching with Inequality in Voluntary Contribution Games. *Comput. Econ.* **2017**, in press.
23. Gunnthorsdottir, A.; Vragov, R.; Shen, J. TACIT Coordination in Contribution-Based Grouping with Two Endowment Levels. *Res. Exp. Econ.* **2010**, *13*, 13–75.
24. Okun, A.M. *Equality and Efficiency—The Big Tradeoff*; The Brookings Institution: Washington, DC, USA, 1975.

25. Harsanyi, J. Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking. *J. Polit. Econ.* **1953**, *61*, 434–435.
26. Binmore, K. *Natural Justice*; Oxford University Press: Oxford, UK, 2005.
27. Harsanyi, J.C.; Selten, R. *A General Theory of Equilibrium Selection in Games*; MIT Press: Cambridge, MA, USA, 1988.
28. Atkinson, A.B. On the measurement of inequality. *J. Econ. Theory* **1970**, *2*, 244–263.
29. Jones-Lee, M.W.; Loomes, G. Discounting and Safety. *Oxf. Econ. Pap.* **1995**, *47*, 501–512.
30. Dorfman, R. A formula for the Gini coefficient. *Rev. Econ. Stat.* **1979**, 146–149.
31. Gini, C. Concentration and dependency ratios. *Riv. Polit. Econ.* **1997**, *87*, 769–792.
32. Blume, L.E. The statistical mechanics of strategic interaction. *Games Econ. Behav.* **1993**, *5*, 387–424.
33. Mäs, M.; Nax, H.H. A behavioral study of “noise” in coordination games. *J. Econ. Theory* **2016**, *162*, 195–208.
34. McKelvey, R.D.; Palfrey, T.R. Quantal Response Equilibria for Normal Form Games. *Games Econ. Behav.* **1995**, *10*, 6–38.
35. Cinyabuguma, M.; Page, T.; Putterman, L. Cooperation under the threat of expulsion in a public goods experiment. *J. Public Econ.* **2005**, *89*, 1421–1435.
36. Charness, G.B.; Yang, C.L. *Endogenous Group Formation and Public Goods Provision: Exclusion, Exit, Mergers, and Redemption*; University of California at Santa Barbara: Santa Barbara, CA, USA. Unpublished work, 2008.
37. Ahn, T.; Isaac, R.M.; Salmon, T.C. Endogenous group formation. *J. Public Econ. Theory* **2008**, *10*, 171–194.
38. Ehrhart, K.; Keser, C. *Mobility and Cooperation: On the Run*; Centre interuniversitaire de recherche en analyse des organisations (CIRANO): Montréal, Canada. Unpublished work, 1999.
39. Coricelli, G.; Fehr, D.; Fellner, G. Partner Selection in Public Goods Experiments. *J. Confl. Resolut.* **2004**, *48*, 356–378.
40. Page, T.; Putterman, L.; Unel, B. Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *Econ. J.* **2005**, *115*, 1032–1053.
41. Brekke, K.; Nyborg, K.; Rege, M. The fear of exclusion: Individual effort when group formation is endogenous. *Scand. J. Econ.* **2007**, *109*, 531–550.
42. Brekke, K.; Hauge, K.; Lind, J.T.; Nyborg, K. Playing with the good guys. A public good game with endogenous group formation. *J. Public Econ.* **2011**, *95*, 1111–1118.
43. Simon, H.A. A mechanism for social selection and successful altruism. *Science* **1990**, *250*, 1665–1668.
44. Bowles, S.; Gintis, H. *A Cooperative Species—Human Reciprocity and Its Evolution*; Princeton University Press: Princeton, NJ, USA, 2011.
45. Fehr, E.; Camerer, C. Social neuroeconomics: The neural circuitry of social preferences. *Trends Cognit. Sci.* **2007**, *11*, 419–427.
46. Allchin, D. *The Evolution of Morality*; Evolution: Education and Outreach; Springer: Berlin, Germany, 2009.
47. Hamilton, W.D. The Genetical Evolution of Social Behaviour I. *J. Theor. Biol.* **1964**, *7*, 1–16.
48. Hamilton, W.D. The Genetical Evolution of Social Behaviour II. *J. Theor. Biol.* **1964**, *7*, 17–52.
49. Nowak, M.A. Five rules for the evolution of cooperation. *Science* **2006**, *314*, 1560–1563.
50. Alger, I.; Weibull, J. Homo Moralis—Preference Evolution Under Incomplete Information and Assortative Matching. *Econometrica* **2013**, *81*, 2269–2302.
51. Grund, T.; Waloszek, C.; Helbing, D. How Natural Selection Can Create Both Self- and Other-Regarding Preferences, and Networked Minds. *Sci. Rep.* **2013**, *3*, 1480.
52. Newton, J. The preferences of Homo Moralis are unstable under evolving assortativity. *Int. J. Game Theory* **2017**, *46*, 583–589.
53. Nax, H.H.; Rigos, A. Assortativity evolving from social dilemmas. *J. Theor. Biology* **2016**, *395*, 194–203.
54. Newton, J. Shared intentions: the evolution of collaboration. *Games Econ. Behav.* **2017**, *104*, 517–534.
55. Cabral, L.M.B. Asymmetric Equilibria in Symmetric Games with Many Players. *Econ. Lett.* **1988**, *27*, 205–208.
56. Gardiner, C.W. *Stochastic Methods*; Springer: Berlin, Germany, 2009.

