Heinrich H. Nax, Ryan O. Murphy & Dirk Helbing

Stability and welfare of 'merit-based' group-matching mechanisms in voluntary contribution game

Working paper

Stability and welfare of 'merit-based' group-matching mechanisms in voluntary contribution games*

Heinrich H. Nax[†], Ryan O. Murphy[‡] & Dirk Helbing[§]

October 21, 2014

Abstract

We study the stability and welfare properties of *merit-based* (meritocratic) group-matching mechanisms in voluntary contribution games. Meritocratic matching in this context means that players tend to be assortatively grouped according to their contributions. We let regimes differ from one another with respect to their matching fidelity. The stability analysis summarizes as follows. When there is not enough meritocracy, the only equilibrium state is universal free-riding. Above a first threshold, several Nash equilibria above free-riding emerge, but only the free-riding equilibrium is stochastically stable. There exists a second meritocratic threshold, above which an equilibrium with high contributions becomes the unique stochastically stable state. This operationalization of meritocracy sheds light on critical transitions, that are enabled by contribution-assortative matching, between equilibria related to "tragedy of the commons" and equilibria with higher expected payoffs for all players. Transitions to the more efficient equilibria come at small inequality costs, so that welfare is typically maximized at the second meritocracy threshold.

JEL classifications: C73, D02, D03, D63

Keywords: meritocracy, voluntary contribution, public goods, stochastic stability, evolutionary stability, assortative matching, welfare, efficiency-equality tradeoff

---

*The authors would like to thank Lukas Bischofberger for help with simulations, Bary Pradelski, Anna Gunnthorsdottir, Michael Mäs, Matthias Leiss, Francis Dennig, Stefan Seifert and Stefano Duca for helpful comments on earlier drafts, Peyton Young for help with framing of the questions, Luis Cabral for help with proposition (12), Ingela Alger and Jörgen Weibull for a helpful discussion, and finally members of GESS at ETH Zurich, the participants at *Norms Actions Games 2014* and at the 25^{th} *International Conference on Game Theory 2014* at Stony Brook for helpful feedback. All remaining errors are ours.

†Corresponding author: Department of Social Sciences, ETH Zürich, Clausiusstrasse 50-C3, 8092 Zürich, Switzerland. hnax@ethz.ch. Nax acknowledges support by the European Commission through the ERC Advanced Investigator Grant 'Momentum' (Grant No. 324247).

‡Department of Social Sciences, ETH Zürich.

§Department of Social Sciences, ETH Zürich. Helbing acknowledges support by the European Commission through the ERC Advanced Investigator Grant 'Momentum' (Grant No. 324247).

1 Introduction

The argument in favor of meritocracy is that meritocratic incentives, such as rewarding effort or performance, promise efficiency gains. In environments such as education, job matching, or marriage markets, however, the downside of meritocracy is that these incentives may exacerbate inequalities (Young 1958, Greenwood et al. 2014). A meritocratic regime in these environments may therefore turn out to be ultimately undesirable from the perspective of a social planner who is averse to inequality (Arrow et al. 2000). In this paper, we study a class of meritocratic regimes in the context of voluntary contributions games where players first commit to individual contribution levels and are then matched based on their decisions. We shall consider regimes in which players are assortatively but imperfectly grouped by their contributions, and shall refer to this as *meritocratic matching*. A version of such a matching mechanism was first introduced in the *group-based mechanism* of Gunnthorsdottir et al. (2010a). In contrast to the standard *voluntary contributions mechanism* (Isaac et al. 1985), where the only equilibrium state is zero contributions, meritocratic matching can enable equilibria with high contribution levels. In this paper, we shall demonstrate that sufficient levels of meritocracy not only enable these new equilibria but also stabilize them in an evolutionary sense, even if the regime is imperfect. Moreover, we shall show that, other than in the aforementioned social contexts, such as education, job matching, or marriage markets, an intermediate but substantial level of meritocracy almost unambiguously maximizes welfare in such public goods situations. An important assumption in our welfare analysis is related to our stability analysis, namely that the social planner selects only amongst the states that are *stochastic stability* (Foster and Young 1990, Young 1993).

Meritocratic matching incentivizes contributing by the promise of grouping contributors with other contributors, and mitigates the incentive to free-ride by the threat of grouping free-riders with other free-riders. This opens a route out of the non-provision impasse, which is the only equilibrium under the standard voluntary contribution mechanism (Isaac et al. 1985).[1] Indeed, Gunnthorsdottir et al. (2010a) show that additional near-efficient equilibria may emerge.[2] In this paper, we shall build on Gunnthorsdottir et al. (2010a)'s mechanism and on the standard voluntary contributions mechanism (Isaac et al. 1985, Isaac and Walker 1988) by making the following three contributions. First, we analyze how robust the predictions concerning equilibria are with respect to imperfections in the meritocratic matching process. Second, we study the stability of the different equilibria including mixed strategy Nash equilibria from an evolutionary game theory perspective. Third, we assess the different candidate equilibria in terms of their efficiency and inequality properties and use these assessments to make welfare comparisons between meritocracy regimes.

The robustness and stability characteristics of meritocratic matching can be summarized in terms of two thresholds. Below a first threshold of meritocratic matching fidelity ("necessary meritocracy"), the only equilibrium state is universal free-riding. Above necessary meritocracy, several Nash equilibria other than free-riding emerge, but only the free-riding equilibrium is

[1]Universal free-riding/ non-provision is related to "tragedy of the commons" in common-pool resource problems (Hardin 1968, Ostrom 1990), Ostrom et al. 1992, but one should not think of them as equivalent due to important psychological differences (Andreoni, 1995b).

[2]We shall refer to the *group-based mechanism* in Gunnthorsdottir et al. (2010a) as "full meritocracy" in our model, because our regimes are all "group-based" in the sense of Gunnthorsdottir et al. (2010a) but vary in the degree of meritocracy.

stochastically stable. Above necessary meritocracy there exists a second threshold ("sufficient meritocracy"), above which an equilibrium with high contributions becomes the unique stochastically stable state. This operationalization of meritocracy sheds light on critical transitions between universal free-riding and equilibria with high contributions.

The comparison of the different equilibria reveals that high contributions equilibria *ex-ante* payoff-dominate (Harsanyi and Selten, 1988) the zero contributions equilibrium. However, there is perfect payoff equality in the latter, while there is payoff heterogeneity in the former. Comparing the high contributions equilibria with the zero contributions equilibrium *ex post*, all free-riders are strictly better-off, and under certain conditions all contributors are also better-off. Hence, an increase of meritocratic matching fidelity at or above sufficient meritocracy under these conditions promises unambiguous welfare gains relative to less meritocratic regimes. Indeed, the social planner maximizes social welfare by setting the meritocratic matching fidelity at sufficient meritocracy. Any level of meritocracy beyond that would increase inequality without increasing efficiency. Any level of meritocracy below that would lead to perfect equality but also to maximum inefficiency.

The remainder of this paper is structured as follows. Next, we discuss the related literature, including related models of public goods provision and the broad conceptual approach. In section 3, we develop a formal model of meritocratic matching, calculate its equilibria, and detail their stability and welfare properties. We conclude in section 4.

2 Related literature

Contributions mechanisms in experimental economics. As mentioned above, the model that is most closely related to ours is the *group-based mechanism* by Gunnthorsdottir et al. (2010a). In fact, the *meritocratic matching* regimes that we propose in this paper fill the space between the standard *voluntary contributions mechanism* with random rematching (Isaac et al. 1985, Andreoni 1988) and the group-based mechanism. Our model can be understood to provide evolutionary and welfare underpinnings for these two models that we shall now detail.

The voluntary contributions mechanism as introduced in Isaac et al. (1985), Isaac and Walker (1988) and as adapted to random group re-matching by Andreoni (1988) is the "no meritocracy" regime in our model. It proceeds as follows. First, players in a finite population make voluntary contributions. Second, they are randomly sorted into a finite number of equal-sized groups. Importantly, players' chosen contributions have no effect on which group they will be part of. Finally, payoffs are realized dependent on the underlying marginal per capita rate of return, which aggregates the contributions in each group, multiples the sum by a coefficient (which is greater than one but smaller than the size of the group), and then evenly divides the product among the group members. With rates of return between one and the group size, total welfare is highest if everybody makes the maximal contribution, but, given any vector of contributions by others, each player maximizes his own payoff by contributing nothing; as a result, free-riding is the unique dominant strategy.

The group-based mechanism by Gunnthorsdottir et al. (2010a) proposes an alternative group-formation regime. Instead of random rematching, players are ranked in order of magnitude of contributions and subsequently matched into groups based on that order. The group-based mechanism represents "full meritocracy" in our model. The ranking is such that no player who contributes less than another is ranked higher than him and possible ties are randomly broken.[3] As in the voluntary contributions mechanism, equal-sized groups form, but now this takes place according to rank. Suppose the group size is s. Then given the ranking, the top s contributors form group one, the next s contributors form group two, etc. Finally, payoffs realize as previously. Under full meritocracy, the free-riding Nash equilibrium continues to exist, but other equilibria may emerge depending on relative group size and rate of return. The focus of the paper by Gunnthorsdottir et al. (2010a) is the near-efficient Nash equilibrium in which a large majority of players contributes everything and only a small fraction of players free-rides, which is also what actually is experimentally observed. This has since been validated in other studies (see Gunnthorsdottir and Thorsteinsson 2010; Gunnthorsdottir et al. 2010b). See also a related experimental investigation of pairwise tournaments à la Becker (1973); Cole et al. (1992) played in this fashion by Rabanal and Rabanal.

More generally, our model and that of Gunnthorsdottir et al. (2010a) contribute to the theoretical underpinnings of group formation in the literature on voluntary contributions mechanisms in experimental economics; see Ledyard 1995 and Chaudhuri 2011 for reviews of that literature. The first formal model of a voluntary contributions game was introduced in Isaac et al. (1985), Isaac and Walker (1988), and this type of mechanism will represent the "no meritocracy" regime in our model. In these models, groups within which public goods are provided remain fixed (Isaac et al., 1985). To disentangle learning from reputation effects in repeated games, this mechanism was extended to random group re-matching (e.g. Andreoni, 1988).[4] The important feature in both fixed and random group matching is that group matching is exogenous, and contributions have no effect on this. An important avenue has been to model how groups form endogenously under different, non-random, mechanisms. Cinyabuguma et al. (2005) and Charness and Yang (2008) consider endogenous group formation and group-size determination via voting. Ehrhart and Keser (1999) and Ahn et al. (2008) study the effects of free group entry and exit. Coricelli et al. (2004) analyze roommate-problem stable matching in pairwise-generated public goods. Page et al. (2005) study rematching in repeated games based on reputation, and Brekke et al. (2007), Brekke et al. (2011) consider the effects of group coordination with signaling. A common feature of the endogenous-group-formation literature is that non-random group formation dynamics can stabilize high contributions equilibria under certain conditions. It is an avenue for future research to study the stability and welfare properties of these other group formation dynamics and under institutional imperfections, as we study here for the group-based mechanism of Gunnthorsdottir et al. (2010a).

Also related to but separate from our approach is the literature on team-based reward schemes in labour market applications (e.g. Dickinson and Isaac 1998; Irlenbusch and Ruchala 2008) where employee team outputs but not individual contributions are observable. In these schemes groups whose aggregate output is above the average are rewarded by bonuses paid to all team members and/ or that groups whose aggregate output is below the average are sanctioned by fines to all team members. It has been shown, theoretically and in the laboratory, that such schemes

[3]Note that this mechanism bears an element of chance due to random tie-breaking, which means that it is possible that 'unlucky' contributors end up in groups with players who contribute less, despite contributing equal amounts to other 'lucky' contributors matched in better groups.

[4]Indeed, most of the literature on learning in public goods games uses variants of Andreoni's random rematching (e.g. Andreoni, 1988, 1993, 1995a; Bayer et al., 2013; Ferraro and Vossler, 2010; Fischbacher and Gaechter, 2010; Goeree et al., 2002; Nax et al., 2013; Palfrey and Prisbrey, 1996, 1997).

provided the employer uses the right 'manipulation' of rewards and sanctions can overcome the free-riding problem in a similar manner to ours. The important difference with our approach is that we generalize the matching procedure, but do not introduce payoff transfers.

Meritocracy and welfare. On a conceptual level, our paper contributes to the literature in political philosophy on meritocratic forms of rule. In political philosophy, "meritocracy" refers to the selection and promotion of individuals (or groups of individuals) based on earned credentials, rather than other principles, such as lineage, luck, looks, or other subjective or arbitrary characteristics that are not directly relevant to assess a person's performance or capacity. Although the term "meritocracy" is relatively recent (Young, 1958), the principle underlying such institutional mechanisms can be traced back to ancient history and has been identified in many independent cultures. Indeed, several institutions of early modern civilizations (e.g. China and Greece) were meritocratic, and meritocratic practice was advocated explicitly by their thinkers (e.g. Confucius, Aristotle, and Plato). Historically, these institutions included the selection of officials and councilmen, reward and promotion schemes, and access to education.[5] Until today, meritocratic institutions like the Chinese civil service examination are in place. Other current examples include honorary circles, bonus wage schemes, etc.

The scope of this paper is limited to a class of public goods games based on voluntary contributions. These games are strategic interactions that are, on the one hand, non-constant sum, and, on the other, group-based rather than individual-based. In other situations, where there is a constant sum of resources and/ or when meritocratic incentives require welfare transfers between individuals, there exists an inherently difficult trade-off between efficiency and equality. Previous research has identified precisely this feature as the main weakness of meritocratic regimes; i.e. that meritocracy increases heterogeneity of payoffs. Indeed, most of the modern political philosophy of meritocracy has focussed on this issue by analysis of situations with efficiency-equality tradeoffs (Young 1958, Arrow et al. 2000). In the often-discussed context of education, for example, it is argued that merit-based rewards first lead to inequality of opportunities and ultimately to growing inequality in wealth. Similarly, an increasing income-assortative matching on the marriage market is shown to contribute to growing income inequality at household levels (Greenwood et al., 2014). As Amartya Sen points out in Arrow et al. (2000), however, the inequality feature of meritocracy is not a general characteristic of meritocratic institutions *per se*, but rather the result of interpreting what constitutes "merit" without distributional concerns in these situations. This distinguishes our study of non-constant-sum situations. Naturally, distributive costs are substantial when meritocracy induces little efficiency gains or even none as in constant-sum environments. In strictly non-constant sum contexts such as contributions games, as we shall consider in this paper, however, meritocratic matching turns out to be a mechanism that is able to promote large efficiency gains at little distributive cost, and under certain conditions even none.

Assortative matching and preference evolution in public goods games. Our paper complements research on cooperative phenomena that arise from non-selfish preferences and altruism (Simon 1990, Bowles and Gintis 2011), in particular in public goods games (e.g. Fehr and Camerer, 2007).[6] It is a well-established finding in evolutionary biology that kin selection

[5]See, for example, Lane (2004) for a description of the reward and promotion scheme in Genghis Khan's army. Another famous example, also in place to the present day, is China's civil service examination (Miyazaki, 1976).

[6]To use the terminology of Allchin (2009), our paper studies a 'system' rather than moral 'acts' or 'intentions' as is the terminology from other papers.

can lead to pro-social behavior in many situations (e.g. Hamilton 1964a, Hamilton 1964b, Nowak 2006). There are recent papers studying the evolution of such preferences under different *assortative matching* mechanisms in public goods games (Alger and Weibull 2013, Grund et al. 2013).[7] These papers focus on the evolution of preferences that are not purely self-regarding in the neoclassic sense (i.e., *homo oeconomicus*), and show that non-selfish/ other-regarding concerns are evolutionarily stable if interactions amongst agents in the population are sufficiently *assortative*. Such studies complement the analysis of biological mechanisms of kin selection in human interactions.

The interesting link with our model is that the phenomena in terms of players' decisions in, for example, Alger and Weibull (2013) could be re-interpreted *as if* players with *homo-oeconomicus* preferences interacted in our kind of meritocratic world without preference assortativity in matching. If one interpreted only the correlation between players' chosen actions and resulting interactions, it is *as if* assortative preference matching generated our kind of meritocratic matching: essentially, contributors and free-riders have a tendency to be rematched with players doing likewise in ours as well as in Alger and Weibull (2013)'s model. However, in Alger and Weibull (2013) this is not driven because players' actions are evaluated by some exogenous institution (as in our model), but due to the population's homophilic matching tendencies. The *as-if* meritocratic matching phenomenon is then a consequence of emergent *non-homo-oeconomicus* preferences and the mixing constraints. As in our model, Alger and Weibull (2013) consider a range of regimes between no assortativity (matching parameter is zero) and full assortativity (matching parameter is one).

The key difference between their model and ours, however, is that if a player was to mutate from a non-selfish player into *homo oeconomicus* again, he would always choose to free-ride. In our model, the *homo oeconomicus* rationally chooses to contribute. Meritocratic matching drives this result. Thus, we complement previous studies by analysis of matching processes that explicitly assort agents based on a meritocratic measure of their chosen actions. In principle, we are agnostic as to the origins of that institution.[8] We simply assume that it exists and study the resulting regime's properties as a function of the meritocratic fidelity. We shall show that sufficient meritocracy implements high levels of stable contributions in a *homo-oeconomicus* population. Meritocratic matching is therefore a mechanism that solves the tragedy of the commons while populated with *homines oeconomici*, without changing the basic payoffs of the game. Players contribute based on purely egoistic and fully rational motivations, without reputation-sensitive concerns, and without hope to be "recognized" for contributing or fear to be "stigmatized" for free-riding (e.g. Andreoni and Petrie 2004, Samek and Sheremeta 2014). Our mechanism therefore works without any non-selfish motivations or non-material incentives. It is an avenue for future research to extend the analysis of social preferences from standard environments (e.g. Fehr and Camerer, 2007; Fehr and Gaechter, 2000, 2002; Fehr and Schmidt, 1999) to our kind of meritocratic matching environment.

[7]In Grund et al. (2013) this is the main focus of the paper, in Alger and Weibull (2013) it is an example of a class of games.

[8]Even though there is some supportive evidence that players endogenously may be able to rank each other in a way that approaches such an exogenous system (Ones and Putterman, 2007).

3 Meritocratic matching in voluntary contributions games

Before we proceed to formalize the set-up of our model, we would like to provide more intuition for the basic flavour of *meritocratic matching.* While none of the following real-world institutions coincides one-to-one with meritocratic matching as it will be instantiated in our model of a simple, linear, symmetric public goods game, all of them mirror meritocratic matching's key features. Entrance examinations to schools or universities, for example, assort individuals based on an imperfect measure of applicants' adequacies for different streams of education and to different schools. An important feature of this sorting mechanism is that the resulting differences in educational quality amongst the different schools are not only determined by the institutional design, but also by the different quality levels of students present in them. Better students tend to study with better students, and worse students with worse students. The incentive to work hard for the examinations is getting into a good school. For an alternative example, imagine an assortative employment regime with team-based payments that rewards employees for performance by matching them with similarly performant employees. Real-world situations with this structure include trading desks in large investment banks, and again this type of competitive grouping incentivizes hard work in order to get into better teams. Finally, we would like to point out the similarity in spirit to the nature of team formation in professional sports, where performant athletes tend to be rewarded by joining successful teams with better contracts. Importantly, all of these real-world examples of meritocratic matching mechanisms are typically both noisy and not always fair. Our model shows how such a mechanism may work, and why a degree of inherent imperfection may indeed be the welfare-optimal mechanism.

3.1 A voluntary contributions game with meritocratic matching

Suppose population $N = \{1, 2, ..., n\}$ plays the following game, of which all aspects are common knowledge. The game is divisible in three steps. First, players make simultaneous voluntary contributions. Second, players receive ranks that imperfectly represent their contributions. Third, groups and payoffs realize based on the ranking.

Step 1. Voluntary contributions

Player $i \in N$ decide simultaneously whether to *contribute* or *free-ride*; we shall write $c_i = 0$ for free-riding and $c_i = 1$ for contributing, yielding the contribution vector $c = \{c_i\}_{i \in N}$. Given some player i, denote by c_{-i} the contribution vector excluding him.

Step 2. Ordering as a function of contributions

An authority imperfectly observes the contribution vector c and/ or imperfectly ranks players according to observed contributions. The measure of ranking precision is given by parameter $\beta \in [0, 1]$. The characteristics of the regimes summarize as follows: *(i) no meritocracy* ($\beta = 0$), all rankings are equally likely, and all players have the same expected rank; *(ii)* in *full meritocracy* ($\beta = 1$), only "perfect" rankings are possible so that all contributors will have a higher rank than all free-riders; *(iii)* in the *intermediate meritocracy* range, when $\beta \in (0, 1)$, all rankings have positive probability, but enough contributors have a higher expected rank than free-riders.

Formally, let $\Pi = \{\pi_1, \pi_2, ..., \pi_{n!}\}$ be the set of *orderings* (permutations) of N. Given any $\pi \in \Pi$, denote by k_i the case when rank $k \in \{1, 2, ..., n\}$ is taken by player $i \in \{1, 2, ..., n\}$. Write $\widehat{\pi}$ for a *perfect* ordering if, for all pairs of players i, j, $k_i < k_j \Rightarrow c_i \geq c_j$, that is, all free-riders are ranked below contributors. Any other ordering is called a *mixed* ordering, and is denoted by $\widetilde{\pi}$ (i.e. at least one free-rider is ranked above a contributor). Given regime $\beta \in [0, 1]$, the probability distribution over orderings, $P(\Pi)$, is a function of β and c, $P(\Pi) = F(c, \beta)$. Write f_π^β for the probability of a particular ordering, $\pi \in \Pi$, under β. Similarly, write f_{ik}^β for the probability that agent i takes rank k given β, and $\overline{k}_i^\beta$ for i's expected rank. We shall write $\overline{k}_i^\beta(c_i)$ to indicate that i's expected rank is a function of his contribution. Finally, define $\mathbf{E}\left[\overline{k}_i^\beta(c_i = 0) - \overline{k}_i^\beta(c_i = 1)\right]$ as the *expected rank difference* from contributing versus free-riding.

We shall assume that all functions f are continuous in β, and that the following properties are the key ingredients to constitute a 'meritocratic matching' mechanism:

(i) no meritocracy. if $\beta = 0$, then, for any c, $f_\pi^0 = 1/n!$ for all $\pi \in \Pi$; hence $\overline{k}_i^\beta = \frac{(n+1)}{2}\ \forall i$

(ii) full meritocracy. if $\beta = 1$, then, for any c with $\sum_{i \in N} c_i = m$, $f_{\widetilde{\pi}}^1 = 0$ for all mixed orderings $\widetilde{\pi}$, and $f_{\widehat{\pi}}^1 = \frac{1}{m!(n-m)!}$ for all perfect orderings $\widehat{\pi}$; hence $k_i^\beta(c_i = 1) = \frac{m+1}{2}$ for all i with $c_i = 1$, and $k_j^\beta(c_j = 0) = \frac{n+m+1}{2}$ for all j with $c_j = 0$

(iii) imperfect meritocracy. if $0 < \beta < 1$, then, for all players i and for any c_{-i},

$$\mathbf{E}\left[\overline{k}_i^\beta(c_i = 0) - \overline{k}_i^\beta(c_i = 1)\right] > 0, \tag{1}$$

$$\partial\mathbf{E}\left[\overline{k}_i^\beta(c_i = 0) - \overline{k}_i^\beta(c_i = 1)\right] / \partial\beta > 0. \tag{2}$$

We shall also require some regularities in terms of the effects of meritocratic matching on expected payoffs (see appendix). Next, we shall discuss several instantiations that satisfy our requirements.

Step 3. Grouping as a function of orderings

Groupings. Finally, groups form based on the ranking and payoffs realize based on the contributions made in each group. Given π, we assume that m groups $\{S_1, S_2, ..., S_m\}$ of a fixed size $s < n$ form the partition ρ of N (where $s = n/m > 1$ for some $s, m \in \mathbf{N}^+$): every group $S_p \in \rho$ (s.t. $p = 1, 2, ..., m$) consists of all players i for whom $k_i \in ((p-1)s + 1, ps]$.

Payoffs. Given contributions c and partition ρ, each $i \in N$ receives $\phi_i(c_i|c_{-i}, \rho)$. Let $\phi = \{\phi_i\}_{i \in N}$ be the payoff vector. When $i \in S$, given the *marginal per capita rate of return* $R := r/s$, i receives

$$\phi_i(c_i|c_{-i}, \rho) = \underbrace{(1 - c_i)}_{\text{remainder from budget}} + \underbrace{(R) * \sum_{j \in S} c_j}_{\text{return from the public good}}. \tag{3}$$

It is standard to assume that $R \in (1/s, 1)$, in which case contributing is socially beneficial under all mechanisms, but a strictly dominated strategy under "no meritocracy" (details are provided in the analysis of the Nash equilibria in the next section).

3.1.1 Meritocratic matching examples

Under meritocratic matching, a player's expected rank difference (expression (1)) is always positive and increasing in β (expression (2)). There are many functional assumptions that satisfy these requirements, one of which is the following:

Meritocratic matching via logit. Given β and c, let $l_i := \frac{\beta c_i}{1-\beta}$. Suppose ranks are assigned according to the following logit-response ordering: if any arbitrary number of $(k-1)$ ranks from 1 to $(k-1) < n$ have been taken by some set of players $S \subset N$ (with $|S| = k-1$), then any player's $i \in \{N \setminus S\}$ probability to take rank k is

$$p_i(k) = \frac{e^{l_i}}{\sum_{j \in N \setminus S} e^{l_j}}. \tag{4}$$

Meritocratic matching via logit is one instantiation of meritocratic matching where the rank ordering is achieved only imperfectly even if the contributions are perfectly observable. Other interpretations of $\beta \in [0,1]$ are *(i)* β represents the fraction of every contributed unit to enter the group-based mechanism and $1-\beta$ to enter the voluntary contributions mechanism, or *(ii)* $(1-\beta)/\beta$ represents some normally distributed noise δ^2 added to the contribution vector c so that contributions are only imperfectly observable, after which the group-based mechanism (with $\beta = 1$) is applied to $x \sim N(c, \delta^2)$.

3.2 Nash equilibria

From expression (3), the expected payoff of contributing c_i given c_{-i} for any i is

$$\underbrace{\mathbf{E}\left[\phi_i(c_i|c_{-i})\right]}_{\text{expected return from } c_i} = \underbrace{1}_{(i)\text{ budget}} - \underbrace{(1-R)*c_i}_{(ii)\text{ sure loss on own contribution}} + \underbrace{R*\mathbf{E}\left[\sum_{j\neq i:\ j\in S_i^\pi} c_j|c_i\right]}_{(iii)\text{ expected return from others' contributions}}, \tag{5}$$

where $S_i^\pi \in \rho$ is the subgroup into which player i is grouped. Note that term (iii), the expected return from others' contributions, is a function of one's own contribution due to meritocratic matching, which, if $c_i = 1$, is increasing in both c_{-i} and β.

First, let us consider candidates for Nash equilibria in pure strategies. Write 1^m for "m players contribute, all others free-ride", and 1^m_{-i} for the same statement excluding player i. The following two conditions must hold for 1^m to constitute a Nash equilibrium:

$$\mathbf{E}\left[\phi_i(1|1^m_{-i})\right] \geq \mathbf{E}\left[\phi_i(0|1^{m-1}_{-i})\right] \tag{6}$$

$$\mathbf{E}\left[\phi_i(0|1^m_{-i})\right] \geq \mathbf{E}\left[\phi_i(1|1^{m+1}_{-i})\right] \tag{7}$$

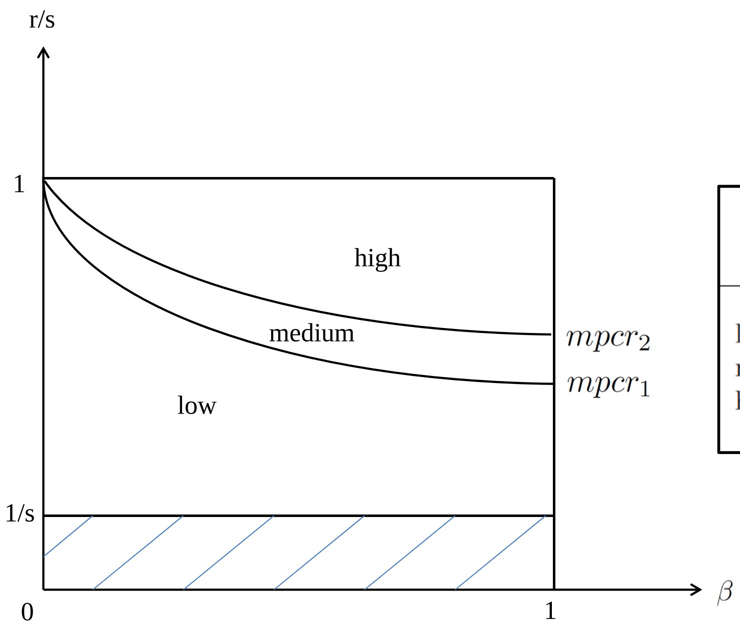
A special case is 1^0 when all players free-ride, and we shall reserve the expression 1^m to refer to cases with $m > 0$. It is easy to verify that 1^0 is always a Nash equilibrium (see Appendix A, proposition 6). Gunnthorsdottir et al. (2010a) show that, when $\beta = 1$, there exists a Nash equilibrium of the form 1^m with $m > 0$ provided $R \geq \frac{n-s+1}{ns-s^2+1} =: \underline{mpcr}$. We shall extend this analysis to show that, given any $R > \underline{mpcr}$, there exists a $\beta < 1$ such that there exists a Nash equilibrium of the form 1^m with $m > 0$. The minimum level of β, denoted by $\underline{\beta}$, for which such a Nash equilibrium exists, is an implicit function that is decreasing in R provided $R > \underline{mpcr}$.

Second, we consider Nash equilibria in symmetric mixed strategies. Write 1^p for "all players contribute with probability $p > 0$", and 1^p_{-i} for the same statement excluding some player i. Again we require $p > 0$ to distinguish from the universal free-riding state. The following condition must hold for 1^p to constitute a Nash equilibrium:

$$\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] = \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]. \tag{8}$$

We shall prove that, for every β, there exists a $R \in (\underline{mpcr}, 1)$ such that there exist two Nash equilibria of the form 1^p with $p > 0$, one with a high $\overline{p}$ and one with a low $\underline{p}$. Write $\overline{mpcr}$ for the necessary marginal per capita rate of return when $\beta = 1$. Expressed differently, given any $R > \overline{mpcr}$, there exists a $\beta < 1$ such that there exist two Nash equilibria of the form 1^p with $\underline{p}$, $\overline{p}$ such that $1 > \overline{p} > \underline{p} > 0$.

The following figure and table qualitatively summarize the existence of different Nash equilibria as a function of R and β (see Appendix A for details).[9] In the table, ✔ indicates that there exists a Nash equilibrium with this structure, ✕ indicates that there exists no Nash equilibrium with this structure:



	1^0	1^m	1^p
low	✔	✕	✕
medium	✔	✔	✕
high	✔	✔	✔

[9] Note that, due to the discreteness of the problem, the lines represented as well-behaved lines (smooth and monotonic), are actually quite different-looking step functions for concrete cases.

Appendix A contains detailed proofs for the existence results. It should be noted that the particular interest of this paper is the analysis of the evolutionary stability and welfare analysis of the system's equilibria as a function of the meritocratic matching parameter β. We shall therefore assume that our implicit bound $\overline{mpcr}$ is satisfied, meaning that all equilibria are at least guaranteed to exist. Thus, our work complements the analysis of Gunnthorsdottir et al. (2010a), where the focus of analysis is the dependence of equilibria existence on the model parameters including the rate of return for the case when $\beta = 1$.

For the case when $R > \overline{mpcr}$, the following observations summarize the arguments from Appendix A:

A. Free-ride trumps contribute (unconditional case):
$\mathbf{E}\left[\phi_i(0|1^m_{-i})\right] > \mathbf{E}\left[\phi_i(1|1^m_{-i})\right]$ for $\beta < \underline{\beta}$ and for any $1_m \geq 0$

Observation A states that, when there is not enough meritocracy, then free-riding is a better reply given any set of actions by the other players.

B1. Free-ride trumps contribute (conditional case).
$\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] > \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]$ for $\beta \geq \underline{\beta}$ and for any $p < \underline{p}$ or $p > \overline{p}$

B2. Contribute trumps free-ride (conditional case).
$\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] < \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]$ for $\beta \geq \underline{\beta}$ and for any $p \in (\underline{p}, \overline{p})$

Observation B states that, when meritocracy is above a "necessary meritocracy" level ($\beta \geq \underline{\beta}$), then contributing is a better reply for intermediate proportions of contributing of contributions (the range of which is given by the symmetric mixed strategy Nash equilibria probabilities), and free-riding is a better reply outside that range.

C. Contribute-free-ride indifference.
$\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] = \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]$ for $\beta \geq \underline{\beta}$ and for $p = \underline{p}$ or $\overline{p}$

Observation C is the condition for a symmetric mixed strategy Nash equilibria to exist. We shall refer to $\overline{p}$ as the "near-efficient" symmetric mixed strategy Nash equilibrium, and to $\underline{p}$ as the "less efficient" symmetric mixed strategy Nash equilibrium.

3.3 Stability

In this section, we shall analyze the stability properties of states in terms of *evolutionary stability* (Maynard Smith and Price, 1973) under replicator dynamics (Helbing, 1996; Taylor and Jonker, 1978; Weibull, 1995) and in terms of *stochastic stability* (Foster and Young, 1990) under constant error rates (Kandori et al., 1993; Young, 1993). The motivation for this analysis is that we view β as a policy choice. We want to understand how the stability of different equilibria depends on the level of meritocracy in matching.

We shall begin by defining the following dynamic game played by agents that we shall assume act myopically. A large population $N = \{1, 2, ..., n\}$ plays our game in continuous time. Let a state of the process be described by p, which is a proportion of players contributing, while the remaining $(1 - p)$ free-ride. Let $\Omega = [0, 1]$ be the state space.

3.3.1 Replicator dynamics

Suppose the two respective population proportions grow according to the following *replicator equation* (Maynard Smith and Price 1973, Taylor and Jonker 1978, Helbing 1996):

$$\partial p/\partial t = (1-p)p\left(\mathbf{E}\left[\phi_i(1|1^p)\right] - \mathbf{E}\left[\phi_i(0|1^p)\right]\right) \tag{9}$$

Evolutionarily stable states. A state where a proportion $\bar{p}$ of players plays $c_i = 1$ is *evolutionarily stable* (ESS) if, for all $p \in [0,1]$ in some arbitrarily small ϵ-neighbourhood around $\bar{p}$, $\partial p/\partial t > 0$ at $p < \bar{p}$, $\partial p/\partial t = 0$ at $p = \bar{p}$, and $\partial p/\partial t < 0$ at $p > \bar{p}$.[10]

Lemma 1. *Given population size n, group size s such that $n > s > 1$ and rate of return r such that $R \in (\overline{mpcr}, 1)$, there exists a $\underline{\beta} > 0$ below which the only ESS is the free-riding Nash equilibrium. When $\beta > \underline{\beta}$, the free-riding Nash equilibrium remains ESS, and, in addition, the population proportions given by the near-efficient symmetric mixed-strategy Nash equilibrium is also an ESS.*

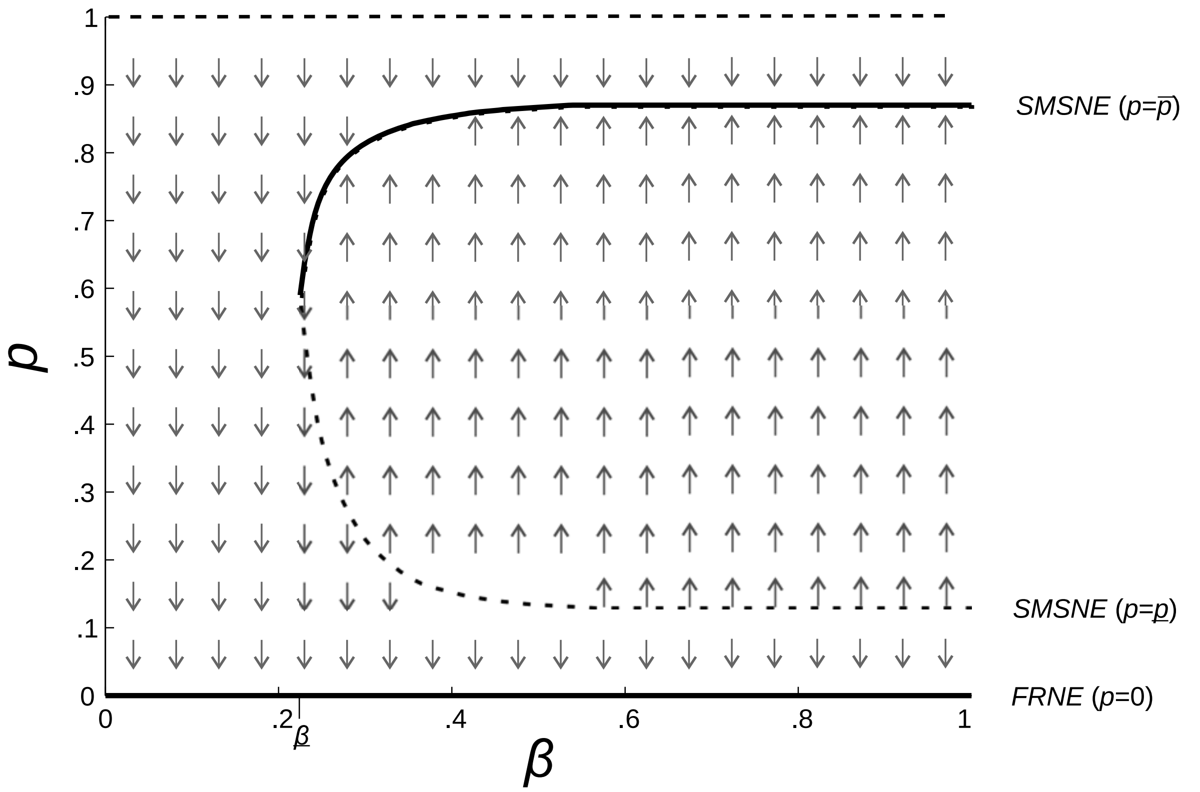
Proof. The proof of Lemma 1 and the cut-off structure of the ESS as given by the analysis of symmetric mixed strategy Nash equilibria in Proposition 7 (see Appendix A for both) led to the summary of best replies as given by Observations A-C. Denote by $\underline{\beta}$ the **necessary meritocracy** level in Proposition 7. Observation A implies that the the only ESS when $\beta < \underline{\beta}$ is given by the free-riding Nash equilibrium because there is only one Nash equilibrium. Observations B1 implies that the free-riding Nash equilibrium is also ESS when $\beta \geq \underline{\beta}$. Observation B1, B2 and C, jointly, imply that population proportions given by the near-efficient symmetric mixed-strategy Nash equilibrium also describe an ESS since it is a local attractor. □

Remark 2. *As replicator dynamics increase the population size (population becomes 'large'), the possible interval for R converges to $(1/s, 1)$ (as is proven in Proposition 12 in Appendix A). In that sense, Lemma 1 is a general observation about the near-efficient symmetric mixed-strategy Nash equilibrium for any rate of return.*

Figure 1 illustrates the implied replicator phase transitions for proportions of players contributing as a function of β under meritocratic matching via logit (Equation 4) for $s = 4$ and $r = 1.6$ starting with $n = 16$ (note that the phase transitions assume the long-run behavior as the population becomes large). In particular, the figure shows how, for large enough values of β, a relatively small 'jump up' is needed starting at the free-riding equilibrium to reach the basin of attraction of the high-contribution equilibrium. By contrast, for low values of β, a small 'draw down' is sufficient to fall out of the high equilibrium into the free-riding equilibrium.

[10] We shall speak of evolutionarily stable 'states' here instead of evolutionarily stable 'strategies' because of the asymmetry of the state.

Figure 1: Evolutionary stability of population strategies for an economy initialized with $s = 4$, $r = 1.6$ and $n = 16$.



In any case when $\beta < \underline{\beta}$, and when $\beta > \underline{\beta}$ then if p is either in excess of the near-efficient symmetric mixed-strategy Nash equilibrium ($p > \overline{p}$) or short of the less-efficient symmetric mixed-strategy Nash equilibrium ($p < \overline{p}$), $\partial p/\partial t < 0$ (replicator tendency is down). When $\beta > \underline{\beta}$ and $\overline{p} > p > \underline{p}$, then $\partial p/\partial t > 0$ (replicator tendency is up). Depending on the location along the bifurcation, the evolutionarily stable states are therefore when either $p = 0$ (free-riding Nash equilibrium) and when p is set according to the near-efficient symmetric mixed-strategy Nash equilibrium ($p = \overline{p}$). Solid lines in the figure indicate stable equilibria, dashed lines indicate unstable equilibria.

3.3.2 Perturbed dynamics

Instead of replicator dynamics, suppose population N remains fixed, but that the dynamics are perturbed by individual errors. Suppose further that individuals are activated by independent Poisson clocks. The distinct times at which one agent becomes active will be called *time steps* $t = 1, 2, ...$ When individual i at time t is activated (the uniqueness of only one agent's activation is guaranteed by the independence of the Poisson clocks), all agents $j \neq i$ continue playing their previous strategy ($c_j^t = c_j^{t-1}$), while i plays a best reply with probability $1 - \epsilon$, but takes the opposite action with probability ϵ. When both actions are best replies, i replies by playing $c_j^t = c_j^{t-1}$ with probability $1 - \epsilon$ and $c_j^t = 1 - c_j^{t-1}$ with probability ϵ.

State. Let a state of the process be defined by $p^t = \frac{1}{n} \sum_{i \in N} c_i^t$.

Let us begin with a couple of observations. First, the perturbed process (when $\epsilon > 0$) is ergodic, that is, it reaches every state from any state with positive probability in finitely many steps (at most n). The process, therefore, has a unique stationary distribution over Ω. Second, for any given level of β, the absorbing states of the unperturbed process (when $\epsilon = 0$) are the various Nash equilibria in pure strategies of the game as identified in section 3.2 (and in particular the free-riding Nash equilibrium and the near-efficient pure-strategy Nash equilibrium).

Stochastically stability. A state p is *stochastically stable* (Foster and Young, 1990) if the stationary distribution as $\epsilon \to 0$ places positive weight on p.

It will be useful to define the "critical mass" necessary to destabilize a given state p.

Critical mass. Let the *critical mass*, $\mathcal{M}_p^\beta \in [0, n-1]$, necessary to destabilize state p given β be the minimum number of players $|S|$ needed to switch strategy simultaneously corresponding to an arbitrary set of players, $S \subset N$, such that as a result of their switch playing current strategy for at least one player in $N \setminus S$ ceases to be a best reply.

Lemma 3. *The stochastically stable state is the near-efficient pure-strategy Nash equilibria if $\mathcal{M}_0^\beta < \mathcal{M}_{\overline{p}}^\beta$, the free-riding Nash equilibrium when $\mathcal{M}_0^\beta > \mathcal{M}_{\overline{p}}^\beta$, and both when $\mathcal{M}_0^\beta = \mathcal{M}_{\overline{p}}^\beta$.*

Proof. When pure strategy Nash equilibria exist, stochastically stable states must be pure strategy Nash equilibria of the unperturbed process. Candidates are the free-riding Nash equilibrium and the $\binom{n}{m}$ near-efficient pure-strategy Nash equilibria.

Obviously, the critical mass for any non-equilibrium state p is $\mathcal{M}_p^\beta = 0$ for all values of β. When $\beta < \underline{\beta}$, there exists no critical mass to destabilize the unique equilibrium which is the free-riding Nash equilibrium; $\mathcal{M}_0^\beta = \emptyset$. In other words, the free-riding Nash equilibrium is the only absorbing state and therefore the unique stochastically stable state. When $\beta = \underline{\beta}$, the near-efficient pure-strategy Nash equilibrium has a critical mass of $\mathcal{M}_{\underline{p}}^\beta = 1$. When $\beta > \underline{\beta}$, for all less-efficient $p \geq \underline{p}$, the critical mass is $\mathcal{M}_{\underline{p}}^\beta = 1$ because one more contribution of some player incentivizes other non-contributors to contribute (see Observations A, B1, B2), or one contribution fewer incentivizes all to not contribute. Moreover, for $\beta > \underline{\beta}$, $\Delta \mathcal{M}_0^\beta / \Delta \beta < 0$ and $\Delta \mathcal{M}_{\overline{p}}^\beta / \Delta \beta > 0$ provided $\Delta \beta$ is large enough. If $\mathcal{M}_{\overline{p}}^1 > \mathcal{M}_0^1$ at $\beta = 1$, then, since $\mathcal{M}_{\underline{p}}^\beta < \mathcal{M}_0^\beta$, it must be that there exists a $\overline{\beta} \in (\underline{\beta}, 1)$ above which the near-efficient pure-strategy Nash equilibrium has a larger critical mass than the free-riding Nash equilibrium.

The proof of the lemma is now a direct application of Theorem 3.1 in Young (1998), and follows from the fact that the resistances of transitions between $p = \overline{p}$ and $p = 0$ are given by the critical masses, thus yielding the stochastic potential for each candidate state. □

3.4 Efficiency and welfare

Outcome. Let (ρ, ϕ) describe an *outcome*, that is, realized groups and payoffs.

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

$$W_e(\phi) = \frac{1}{n(1-e)} \sum_{i \in N} \phi_i^{1-e} \tag{10}$$

When $e = 1$, it is standard that $W_1(\phi) = \frac{1}{n} \prod_{i \in N} \phi_i$, i.e. be the Nash product. Expression (10) is a variant of the social welfare function introduced by Atkinson (1970). It nests both the **Utilitarian** (Benthiam) and **Rawlsian** social welfare functions.[11] When $e = 0$, expression (10) reduces to $W_0(\phi) = \frac{1}{n} \sum_{i \in N} \phi_i$, i.e. a Utilitarian social welfare function measuring the state's efficiency. When $e \to \infty$, expression (10) approaches $W_\infty(\phi) = \min(\phi_i)$, i.e. a Rawlsian social welfare function measuring the state's worst-off utility. Obviously, a Utilitarian social planner prefers the near-efficient pure-strategy Nash equilibrium to the free-riding equilibrium. The *ex-post* Rawlsian-optimal equilibrium, however, could be the free-riding Nash equilibrium with perfect equality of payoffs (equal to one for every player) if any player in the near-efficient pure strategy Nash equilibrium receives a payoff of less than one. **Harsanyi**'s social welfare approach (Harsanyi, 1953), on the other hand, would always prefer the near-efficient pure-strategy Nash equilibrium if every contributor and every free-rider is in expectation (i.e. *ex ante*) better-off.[12]

Which equilibrium is preferable in terms of social welfare for any given social welfare function depends on the social planner's relative weights on efficiency and equality and is related to whether an *ex ante* or an *ex post* view is taken with regards to payoff dominance (Harsanyi and Selten, 1988).[13] Critical for this assessment is the inequality aversion e. For the economy illustrated in Table 1 (with $n = 16$, $s = 4$ and $r = 1.6$), suppose a social planner considers moving from $\beta = 0$ to $\beta = 1$. To assess this, he makes an *ex-post* W_e-comparison. It turns out that for any W_e with $e < 10.3$ he prefers the near-efficient pure-strategy Nash equilibrium, while for a W_e with $e \geq 10.3$ he prefers the free-riding Nash equilibrium.[14]

Welfare assessment assumption. Suppose the social planner sets $\beta \in [0, 1]$ so as to maximize $\mathbf{E}[W_e(\phi)]$, where ϕ are expected to be realizations of stochastically stable states.

[11] See, for example, Jones-Lee and Loomes (1995) for a discussion of this generalization.

[12] Harsanyi's social welfare function is $W_H(\phi) = \frac{1}{n} \sum_{i \in N} \mathbf{E}[\phi_i]$. See, for example, Binmore (2005) for a discussion of the Rawlsian and Harsanyi's 'original position' approach.

[13] ϕ *payoff-dominates* ϕ' if $\phi_i \geq \phi'_i$ for all i, and there exists a j such that $\phi_j > \phi'_j$.

[14] With $e = 10.3$, W_e requires efficiency gains of more than twice the amount lost by any player to compenstate for the additional inequality.

Moreover, assume the social planner expects the near-efficient pure strategy Nash equilibrium (here denoted by $\overline{p}$) to be played when $\mathcal{M}_0^\beta = \mathcal{M}_{\overline{p}}^\beta$ (both are stochastically stable).

Proposition 4. *For any $R > \max\{\overline{mpcr}, 1/(s-1)\}$, there exists a population size $n < \infty$ such that $\mathbf{E}[W_e(\phi); \beta = \overline{\beta}] > \mathbf{E}[W_e(\phi); \beta]$ at "sufficient meritocracy" ($\beta = \overline{\beta}$) for all $\beta \neq \overline{\beta}$ given any parameter of inequality aversion $e \in [0, \infty)$.*

Proof. Suppose there exists a $\overline{\beta} \in (\underline{\beta}, 1)$ above which the near-efficient pure-strategy Nash equilibrium is stochastically stable. Write q_1^n for the probability of having more than one free-rider in any group for a realized outcome (ρ, ϕ) 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 $\mathbf{E}[W_e(\phi); \beta] > (1 - q_1^n) \times W_e(\phi_i = (s-1)R \; \forall i)$. Because, given any $\beta < 1$, $\partial q_1^n / \partial n < 0$, there therefore exists $n < \infty$ above which $\mathbf{E}[W_e(\phi)] > W_e(\phi_i = 1 \; \forall i)$. □

Remark 5. $\mathbf{E}[W_e(\phi); \beta = \overline{\beta}] > \mathbf{E}[W_e(\phi); \beta]$ *at "sufficient meritocracy" ($\beta = \overline{\beta}$) is also the case for n smaller than implied by the proposition when (a) e is set below some bound $e < \infty$ and/ or (b) set above some bound $R > 1/(s-1)$.*

Table 1: Stem-and-leaf plot of individual payoffs for the free-riding Nash equilibrium when $\beta = 0$ and for the near-efficient pure-strategy Nash equilibrium when $\beta = 1$ with $n = 16$, $s = 4$, $r = 1.6$ and $\beta = 1$.

near-efficient pure-strategy NE when $\beta = 1$	payoff	free-riding NE when $\beta = 0$
0	0.0	0
0	0.2	0
0	0.4	0
0	0.6	0
13 14 ($c_i = 1$) 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
1 2 3 4 5 6 7 8 9 10 11 12 ($c_i = 1$) 12	1.6	0
15 16 ($c_i = 0$) 2	1.8	0
24.4	efficiency	16

The stem of the table are 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. Note that the near-efficient pure-strategy Nash equilibrium is more efficient, whereas the free-riding Nash equilibrium is more equitable.

4 Conclusion

In this paper, we have addressed three issues. First, we assessed the robustness of equilibrium predictions for meritocracy levels in between "no meritocracy" and "full meritocracy". We found that the minimum meritocracy threshold ("necessary meritocracy") necessary to enable equilibria with high contributions decreases with the rate of return in excess of the minimum rate of return threshold, below which not even "full meritocracy" enables such equilibria. Second, we analyzed the stability properties of the equilibria. It turned out that there exists a second threshold ("sufficient meritocracy") between "necessary meritocracy" and "full meritocracy" at which both the equilibrium with zero contributions and the near-efficient equilibria are stochastically stable. Above "sufficient meritocracy", the high contributions equilibria are stable, below the zero contributions equilibrium is stable. Finally, we assessed the relative welfare properties of the candidate equilibria and used this to identify the welfare-maximizing meritocracy regime. We demonstrated that setting the meritocracy at "sufficient meritocracy" maximizes welfare for any inequality-averse social welfare functions when the population is large enough. For smaller populations, this is also true when (a) the inequality aversion is not extreme and (b) the rate of return is high. Otherwise, meritocracy should be set to zero, in which case equality is achieved perfectly but all contributions will be zero in equilibrium.

References

T. Ahn, R. M. Isaac, and T. C. Salmon. Endogenous group formation. *Journal of Public Economic Theory*, 10:171–194, 2008.

I. Alger and J.W. Weibull. Homo moralis - preference evolution under incomplete information and assortative matching. *Econometrica*, 81:2269–2302, 2013.

D. Allchin. *The Evolution of Morality*. Evolution: Education and Outreach. Springer, 2009.

J. Andreoni. Why free ride? strategies and learning in public goods experiments. *Journal of Public Economics*, 37:291–304, 1988.

J. Andreoni. An experimental test of the public goods crowding-out hypothesis. *American Economic Review*, 83:1317–1327, 1993.

J. Andreoni. Cooperation in public-goods experiments: kindness or confusion? *American Economic Review*, 85:891–904, 1995a.

J. Andreoni. Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments. *Quarterly Journal of Economics*, 110:1–21, 1995b.

J. Andreoni and R. Petrie. Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics*, 88:1605–1623, 2004.

K. Arrow, S. Bowles, and S. Durlauf. *Meritocracy and Economic Inequality*. Princeton University Press, 2000.

A. B. Atkinson. On the measurement of inequality. *Journal of Economic Theory*, 2:244–263, 1970.

R.-C. Bayer, E. Renner, and R. Sausgruber. Confusion and learning in the voluntary contributions game. *Experimental Economics*, 16:478–496, 2013.

G. S. Becker. A theory of marriage: Part 1. *Journal of Political Economy*, 81:813–846, 1973.

K. Binmore. *Natural Justice*. Oxford University Press, 2005.

S. Bowles and H. Gintis. *A cooperative specieshuman reciprocity and its evolution*. Princeton University Press, 2011.

K. Brekke, K. Nyborg, and M. Rege. The fear of exclusion: individual effort when group formation is endogenous. *Scandinavian Journal of Economics*, 109:531–550, 2007.

K. Brekke, K. Hauge, J. T. Lind, and K. Nyborg. Playing with the good guys. a public good game with endogenous group formation. *Journal of Public Economics*, 95:1111–1118, 2011.

L. M. B. Cabral. Asymmetric equilibria in symmetric games with many players. *Economic Letters*, 27:205–208, 1988.

G. B. Charness and C.-L. Yang. Endogenous group formation and public goods provision: Exclusion, exit, mergers, and redemption. *University of California at Santa Barbara, Economics WP*, 2008.

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

M. Cinyabuguma, T. Page, and L. Putterman. Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*, 89:1421–1435, 2005.

H. J. Cole, G. Mailath, and A. Postlewaite. Social norms, savings behavior, and growth. *Journal of Political Economy*, 100:1092–1125, 1992.

G. Coricelli, D. Fehr, and G. Fellner. Partner selection in public goods experiments. *Economics Series*, 151, 2004.

D. L. Dickinson and R. M. Isaac. Absolute and relative rewards for individuals in team production. *Managerial and Decision Economics*, 19:299–310, 1998.

K. Ehrhart and C. Keser. Mobility and cooperation: On the run. *CIRANO WP*, 99(s-24), 1999.

E. Fehr and C. Camerer. Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences*, 11:419–427, 2007.

E. Fehr and S. Gaechter. Cooperation and punishment in public goods experiments. *American Economic Review*, 90:980–994, 2000.

E. Fehr and S. Gaechter. Altruistic punishment in humans. *Nature*, 415:137–140, 2002.

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

P. J. Ferraro and C. A. Vossler. The source and significance of confusion in public goods experiments. *The B.E. Journal in Economic Analysis and Policy*, 10:53, 2010.

U. Fischbacher and S. Gaechter. Social preferences, beliefs, and the dynamics of free riding in public good experiments. *American Economic Review*, 100:541–556, 2010.

D. Foster and H. P. Young. Stochastic evolutionary game dynamics. *Theoretical Population Biology*, 38:219–232, 1990.

J. K. Goeree, C. A. Holt, and S. K. Laury. Private costs and public benefits: Unraveling the effects of altruism and noisy behavior. *Journal of Public Economics*, 83:255–276, 2002.

J. Greenwood, N. Guner, G. Kocharkov, and C. Santos. Marry your like: Assortative mating and income inequality. *NBER WP*, 19829, 2014.

T. Grund, C. Waloszek, and D. Helbing. How natural selection can create both self- and other-regarding preferences, and networked minds. *Scientific Reports*, 3:1480, 2013.

A. Gunnthorsdottir and P. Thorsteinsson. Tacit coordination and equilibrium selection in a merit-based grouping mechanism: A cross-cultural validation study. *Department of Economics WP*, 0, 2010.

A. Gunnthorsdottir, R. Vragov, S. Seifert, and K. McCabe. Near-efficient equilibria in contribution-based competitive grouping. *Journal of Public Economics*, 94:987–994, 2010a.

A. Gunnthorsdottir, R. Vragov, and J. Shen. Tacit coordinationin contribution-based grouping with two endowment levels. *Research in Experimental Economics*, 13:13–75, 2010b.

W. D. Hamilton. The genetical evolution of social behaviour i. *Journal of Theoretical Biology*, 7:1–16, 1964a.

W. D. Hamilton. The genetical evolution of social behaviour ii. *Journal of Theoretical Biology*, 7:17–52, 1964b.

G. Hardin. The tragedy of the commons. *Science*, 162:1243–1248, 1968.

J. Harsanyi. Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, 61:434–435, 1953.

J. C. Harsanyi and R. Selten. *A General Theory of Equilibrium Selection in Games*. MIT Press, 1988.

D. Helbing. A stochastic behavioral model and a 'microscopic' foundation of evolutionary game theory. *Theory and Decision*, 40:149–179, 1996.

B. Irlenbusch and G. Ruchala. Relative Rewards within Team-Based Compensation. *Labour Economics*, 15:141–167, 2008.

M. Isaac and J. Walker. Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics*, 103:179–199, 1988.

M. R. Isaac, K. F. McCue, and C. R. Plott. Public goods provision in an experimental environment. *Journal of Public Economics*, 26:51–74, 1985.

M. W. Jones-Lee and G. Loomes. Discounting and safety. *Oxford Economic Papers*, 47:501–512, 1995.

M. Kandori, G. J. Mailath, and R. Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61:29–56, 1993.

G. Lane. *Genghis Khan and Mongol Rule*. Greenwood, 2004.

J. O. Ledyard. Public goods: A survey of experimental research. *in J. H. Kagel and A. E. Roth (Eds.), Handbook of experimental economics*, 37:111–194, 1995.

J. Maynard Smith and G. R. Price. The logic of animal conflict. *Nature*, 246:15–18, 1973.

I. Miyazaki. *China's Examination Hell: The Civil Service Examinations of Imperial China*. Weatherhill, 1976.

H. H. Nax, M. N. Burton-Chellew, S. A. West, and H. P. Young. Learning in a black box. *Department of Economics WP, University of Oxford*, 653, 2013.

M. A. Nowak. Five rules for the evolution of cooperation. *Science*, 314:1560–1563, 2006.

U. Ones and L. Putterman. The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *Journal of Economic Behavior and Organization*, 62:495–521, 2007.

E. Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990.

E. Ostrom, J. Walker, and R. Gardner. Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86:404–417, 1992.

T. Page, L. Putterman, and B. Unel. Voluntary association in public goods experiments: reciprocity,mimicryand efficiency. *The Economic Journal*, 115:1032–1053, 2005.

T. R. Palfrey and J. E. Prisbrey. Altruism, reputation and noise in linear public goods experiments. *Journal of Public Economics*, 61:409–427, 1996.

T. R. Palfrey and J. E. Prisbrey. Anomalous behavior in public goods experiments: how much and why? *American Economic Review*, 87:829–846, 1997.

J. P. Rabanal and O. A. Rabanal. Efcient investment via assortative matching: A laboratory experiment. *mimeo.*

S. Samek and R. Sheremeta. Visibility of contributors: An experiment on public goods. *Experimental Economics*, 2014.

H. A. Simon. A mechanism for social selection and successful altruism. *Science*, 250:1665–1668, 1990.

P. D. Taylor and L. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Bioscience*, 40:145156, 1978.

J. Weibull. *Evolutionary Game Theory.* The MIT Press, 1995.

H. P. Young. The evolution of conventions. *Econometrica*, 61:57–84, 1993.

H. P. Young. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions.* Princeton University Press, 1998.

M. Young. *The Rise of the Meritocracy, 1870-2033: An Essay on Education and Equality.* Transaction Publishers, 1958.

Appendices

Appendix A: Nash equilibria

Proposition 6. *For any population size $n > s$, group size $s > 1$, rate of return $r \in (1, s)$, and meritocratic matching factor $\beta \in [0, 1]$, there always exists a* ***free-riding Nash equilibrium*** *such that all players free-ride.*

The proof of Proposition 6 follows from the fact that, given any β and for c_{-i} such that $\sum_{j \neq i} c_j = 0$, we have:

$$1 = \mathbf{E}\left[\phi_i(0|c_{-i})\right] > \mathbf{E}\left[\phi_i(1|c_{-i})\right] = R. \tag{11}$$

Equation 11, in words, means that it is never a best response to be the only contributor for any level of β. If, for any level of β, given any c_{-i}, $\mathbf{E}\left[\phi_i(0)|c_{-i}\right] > \mathbf{E}\left[\phi_i(1)|c_{-i}\right]$ holds for all i, then we have a situation where free-riding is the strictly dominant strategy. In that case, for any level of meritocracy (β), universal free-riding is the unique Nash equilibrium. We shall proceed to show that this is not the case if the marginal per capita rate of return (R) and the meritocratic matching fidelity (β) are high enough.

Recall that 1^m stands for "m players contribute, all others free-ride", 1^m_{-i} for the same statement excluding player i, and that $\underline{mpcr} = \frac{n-s+1}{ns-s^2+1}$.

Proposition 7. *Given population size $n > s$, group size $s > 1$ and rate of return r such that $R \in (\underline{mpcr}, 1)$, there exists a* ***necessary meritocracy*** *level, $\underline{\beta} \in (0, 1)$, above which there is a* ***pure-strategy Nash equilibrium****, where $m > 0$ agents contribute and the remaining $n - m$ agents free-ride.*

Proof. The following two conditions must hold for Proposition 7 to be true:

$$\mathbf{E}\left[\phi_i(1|1^m_{-i})\right] \geq \mathbf{E}\left[\phi_i(0|1^{m-1}_{-i})\right] \tag{12}$$

$$\mathbf{E}\left[\phi_i(0|1^m_{-i})\right] \geq \mathbf{E}\left[\phi_i(1|1^{m+1}_{-i})\right] \tag{13}$$

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 $R \geq \overline{mpcr}$ follows from Theorem 1 in Gunnthorsdottir et al. (2010a), in which case both equations (12) and (13) are strictly satisfied.

The fixed point argument behind that result becomes clear by inspection of terms *(ii)* and *(iii)* in expression (5): 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 $R \in (\underline{mpcr}, 1)$ in which case equations (12) and (13) are strictly satisfied. Note that β has a positive effect on the expected payoff of contributing and a negative effect on the expected payoff of free-riding:

$$\partial \mathbf{E}\left[\phi_i(1|1^m_{-i})\right]/\partial\beta > 0 \quad (14)$$

$$\partial \mathbf{E}\left[\phi_i(0|1^m_{-i})\right]/\partial\beta < 0 \quad (15)$$

When $\beta = 0$, we know that $\phi_i(1|1^m_{-i}) = R < \phi_i(0|1^m_{-i}) = 1$ for any m. However, by existence of the equilibrium with $m > 0$ contributors when $\beta = 1$, provided that $R > \underline{mpcr}$ is satisfied, there must exist some maximum value of $\underline{\beta} \in (0, 1)$, at which either equation $\overline{(12)}$ or equation (13) first binds due to continuity of expressions (14) and (15) in β. That level is the bound on β above which the pure-strategy Nash equilibrium with $m > 0$ exists. □

Remark 8. *Note that, for a finite population of size n, a group size s larger than one implies that $\underline{mpcr} > 1/s$ for Proposition 7 to be true, but as $n \to \infty$, $\underline{mpcr}$ converges to $1/s$.*[15]

A special case of a pure-strategy Nash equilibrium is the **near-efficient pure-strategy Nash equilibrium** (see Gunnthorsdottir et al., 2010a): in our set-up, the near-efficient pure-strategy Nash equilibrium generalizes to the the pure-strategy Nash equilibrium in which m is chosen to be the largest value given n, s, r for which equations (12) and (13) hold. For that m to be larger than zero β needs to be larger than $\underline{\beta}$ (Proposition (7)).

Now we shall compare the asymmetric equilibria in pure strategies (in particular the near-efficient pure-strategy Nash equilibrium) with 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^p_{-i} for the same statement excluding some player i.

In section 3.1, we characterized the key ingredients regarding what constitutes a meritocratic ranking/ matching mechanism in our setting. We shall also require, for sufficiently large n, the case for which the subsequent mixed-strategy arguments apply (replicator equations), the following regularity effects of β on expected payoffs. The ranking/ matching methods that we propose in section 3.1.1 in particular satisfy these assumptions, but the class is more general. Recall that we assumed all functions f to be continuous in β. Moreover we assume the following two regularities:

$\partial \mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial p \geq 0$ and $\partial \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]/\partial p \geq 0$.

$\partial^2 \mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial p^2 \geq 0$ and $\partial^2 \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]/\partial p^2 \leq 0$.

Proposition 9. *Given population size $n > s$ and group size $s > 1$, there exists a rate of return r such that $R \in [\underline{mpcr}, 1)$ beyond which there exists a **necessary meritocracy** level, $\underline{\beta} \in (0, 1)$, 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 **symmetric mixed-strategy Nash equilibrium**. One will have a high $\overline{p}$ (the* near-efficient *symmetric mixed-*

[15] It is easy to check that $\lim_{n\to\infty} \underline{mpcr} = 1/s$.

strategy Nash equilibrium) and one will have a low $\underline{p}$ (the less-efficient *symmetric mixed-strategy Nash equilibrium).*

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

$$\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] = \mathbf{E}\left[\phi_i(1|1^p_{-i})\right], \tag{16}$$

because, in that case, player i has a best response also playing $p_i = p$, guaranteeing that 1^p is a Nash equilibrium. Proposition 7 implies that, if $R > \underline{mpcr}$, equations (12) and (13) are strictly satisfied when $\beta = 1$ for m contributors corresponding to the near-efficient pure-strategy Nash equilibrium. Indeed, expressions (12) and (13) imply lower and upper bounds (see Gunnthorsdottir et al. 2010a) on the number of free-riders given by

$$l = \frac{n - nR}{1 - R + nR - r}, \qquad u = 1 + \frac{n - nR}{1 - R + nR - r}. \tag{17}$$

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 $R \to 1$; (ii) possibly none when $R = \underline{mpcr}$; and (iii) a continuity in R such that there is some intermediate value of $R \in [\underline{mpcr}, 1)$ above which at least one symmetric mixed-strategy Nash equilibrium exists but not below.

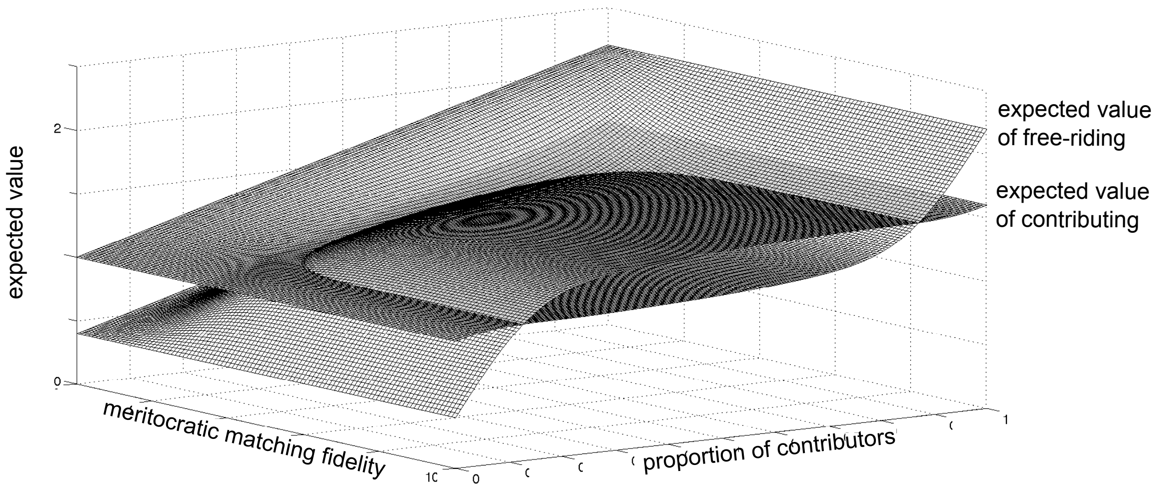
(i) Because $\partial\mathbf{E}\left[\phi_i(c_i|1^p_{-i})\right]/\partial p > 0$ for all c_i, there exists a $p \in (\frac{m-1}{n}, \frac{m+1}{n})$ such that expression (16) holds if $R \to 1$. This is the standard symmetric mixed-strategy Nash equilibrium, which always exists in a symmetric two-action n-person game where the only pure-strategy equilibria are asymmetric and of the same kind as the near-efficient pure-strategy Nash equilibrium (see the proof of Theorem 1 in Cabral 1988). In this case, the presence of the free-riding Nash equilibrium makes no difference because the incentive to free-ride vanishes as $R \to 1$.

(ii) If $R = \underline{mpcr}$, one or both of the equations, (12) or (13), bind. Hence, unless expression (16) holds exactly at $p = m/n$ (which is a limiting case in n that we will address in proposition 12), there may not exist any p such that expression (16) 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 free-riding 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) $\partial\mathbf{E}\left[\phi_i(c_i|1^p_{-i})\right]/\partial r$ 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 (16). 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 $\beta = 1$, $\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] = \mathbf{E}\left[\phi_i(1|1^p_{-i})\right] > 1$. Because of this, a similar argument as in Proposition 7 applies to ensure the existence of some $\underline{\beta} \in (0,1)$ above which the symmetric mixed-strategy Nash equilibrium continues to exist when $R > \overline{mpcr}$: because, at $\beta = 1$, equations (12) and (13) are strictly satisfied and $\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] = \mathbf{E}\left[\phi_i(1|1^p_{-i})\right] > 1$, there therefore must exist some

Figure 2: Expected payoffs of contributing versus free-riding if all others play p and the meritocratic matching fidelity is β for the economy with $n = 16$, $s = 4$, $r = 1.6$.



Expected values *of* $\phi_i(0|1^p_{-i})$ *and* $\phi_i(1|1^p_{-i})$ *are plotted as functions of* probability p *and* meritocratic matching fidelity β *under* meritocratic matching via logit *(equation* 4*). The two planes intersect at the bifurcating symmetric mixed-strategy Nash equilibrium-values of* $\overline{p}$ *and* $\underline{p}$ *(see Proposition* 9*). Notice that the expected values of both actions increase linearly in p when the meritocratic matching fidelity is zero but gain in curvature for larger values, until they intersect at* $\overline{p}$ *and* $\underline{p}$*. Note that* Figure 1 *is a birds-eye view of this figure.*

$\beta < 1$ and $p' < p$ satisfying equation (16) while still satisfying $\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] = \mathbf{E}\left[\phi_i(1|1^p_{-i})\right] > 1$. Note that this implicit bounds here may be different from that in Proposition 7.

Part 2. If $R > \overline{mpcr}$ and $\beta > \underline{\beta}$, existence of two equilibria with $\overline{p} > \underline{p} > 0$ is shown by analysis of the comparative statics of equation (16).

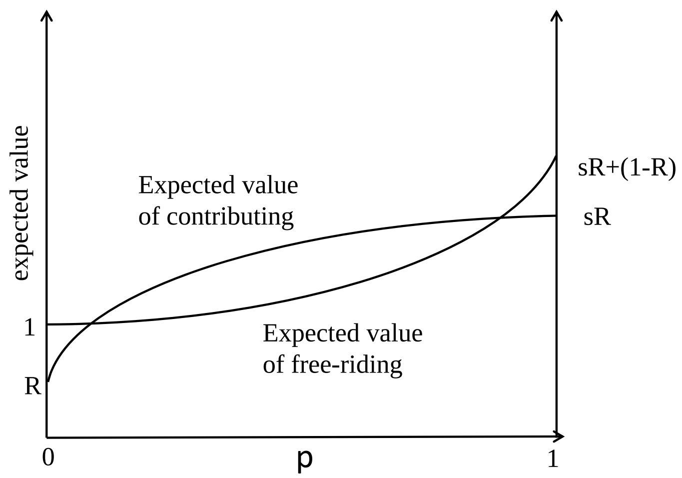
First note that, for any $R > \overline{mpcr}$ and $\beta > \underline{\beta}$, $\partial\mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial\beta < 0$ while $\partial\mathbf{E}\left[\phi_i(1|1^p_{-i})\right]/\partial\beta > 0$. p therefore has to take different values for equation (16) to hold for two different values of β above $\underline{\beta}$. Unclear is whether it has to take a higher or lower value. Note also that both $\partial\mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial p > 0$ and $\partial\mathbf{E}\left[\phi_i(1|1^p_{-i})\right]/\partial p > 0$ for all $\beta \in (0, 1)$. We can rearrange the partial derivative with respect to β of Expression 16, and obtain

$$\partial p/\partial\beta = \frac{\partial\mathbf{E}\left[\phi_i(1|1^p_{-i})\right]/\partial\beta - \partial\mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial\beta}{\partial\mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial p - \partial\mathbf{E}\left[\phi_i(1|1^p_{-i})\right]/\partial p}. \tag{18}$$

Expression 18 is negative if the denominator is negative, because the numerator is always positive.

Observation 10. *The denominator of Equation* 18 *is negative when p is low, and positive when p is high.*

Figure 3: Expected payoffs of contributing versus free-riding if all others play p and the meritocratic matching fidelity is $\beta > \underline{\beta}$.



Expected values *of* $\phi_i(0|1^p_{-i})$ *and* $\phi_i(1|1^p_{-i})$ *are plotted as functions of* probability p *for some fixed* $\beta > \underline{\beta}$*. The two planes intersect at the bifurcating symmetric mixed-strategy Nash equilibrium-values of* $\overline{p}$ *and* $\underline{p}$ *(see Proposition* 9*). The relative slopes of the two curves illustrate the proposition. Note that this figure is a slice through Figure 2 along a value of* $\beta > \underline{\beta}$*.*

First, note that the decision to contribute/ free-ride imply the following expected values at the boundaries: $\mathbf{E}\left[\overline{\phi}_i(0|1^p_{-i})\right] = 1$ when $1^p_{-i} = 0$ and $\mathbf{E}\left[\overline{\phi}_i(0|1^p_{-i})\right] = s * R + (1 - R)$ when $1^p_{-i} = 1$; $\mathbf{E}\left[\overline{\phi}_i(1|1^p_{-i})\right] = R$ when $1^p_{-i} = 0$ and $\mathbf{E}\left[\overline{\phi}_i(1|1^p_{-i})\right] = s * R$ when $1^p_{-i} = 1$. (Figure 3 illustrates.) Due to our continuity assumptions on β and the assumed regularities in terms of the effects of p, we know that $\partial\mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial p \leq \partial\mathbf{E}\left[\phi_i(1|1^p_{-i})\right]/\partial p$ when p is low, and $\partial\mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial p \geq \partial\mathbf{E}\left[\phi_i(1|1^p_{-i})\right]/\partial p$ when p is high. Moreover, notice that the existence of the pure-strategy Nash equilibrium with high contribution for high levels of β ensures that $\mathbf{E}\left[\phi_i(0|1^p_{-i})\right]$ is not always larger than $\mathbf{E}\left[\phi_i(1|1^p_{-i})\right]$. It therefore follows from continuity in β that the RHS of the denominator exceeds the LHS of the denominator when p is low, and that the LSE of the denominator exceeds the RHS of the denominator when p is high, hence the denominator of Equation 18 is negative when p is low, and positive when p is high.

□

Remark 11. *Note that the* ***necessary meritocracy*** *level* $\underline{\beta}$ *in Propositions* 7 *and* 9 *need not be the same. We shall write* $\underline{\beta}$ *for whichever level is larger.*

Proposition 12. *Given group size* $s > 1$*, then, if* $\beta = 1$*, as* $n \to \infty$ *(i)* $1^m/n$ *of the near-efficient pure-strategy Nash equilibrium and* $\overline{p}$ *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 $R > \overline{mpcr}$, i.e. that both symmetric mixed-strategy Nash equilibrium and near-efficient pure-strategy Nash equilibrium exist. Let 1^m describe the near-efficient pure-strategy Nash equilibrium and 1^p describe the near-efficient symmetric mixed-strategy Nash equilibrium. Recall that expressions under (17) summarize the lower and upper bound on the number of free-riders, $(n-m)$ in the near-efficient pure-strategy Nash equilibrium. Taking $\lim_{n\to\infty}$ 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 $R \to 1$.) We know that, if there is one more free-rider than given by the upper bound, then equation (13) is violated. Similarly, if there is one fewer free-rider than given by the lower bound, then equation (12) is violated.

With respect to the near-efficient symmetric mixed-strategy Nash equilibrium, recall that Expression 16 must hold; i.e. $\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] = \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]$. We can rewrite $\mathbf{E}\left[\phi_i(c_i|1^p_{-i})\right]$ as $\mathbf{E}\left[\phi_i(c_i|B)\right]$, 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 $\mathbf{E}[B] = np$ and variance $\mathbf{V}[B] = np(1-p)$. As $n \to \infty$, by the law of large numbers, we can use the same bounds obtained for the near-efficient pure-strategy Nash equilibrium to bound $(B/n) \in [(n-u)/n, (n-l)/n]$, which converges to the unique p at which expression (16) actually holds.[16]

Suppose all players contribute with probability p corresponding to the near-efficient symmetric mixed-strategy Nash equilibrium limit value. Then, $\lim_{n\to\infty}\mathbf{V}[(B/n)] = \lim_{n\to\infty}\frac{p(1-p)}{n} = 0$ for the actual proportion of contributors. Hence, the limit for the range over R necessary to ensure existence converges to that of the near-efficient pure-strategy Nash equilibrium, which by Remark 8 is $(1/s, 1)$. □

Remark 13. *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 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.*

[16]Details concerning the use of the law of large numbers can be followed based on the proof in Cabral (1988).