Meritocratic matching stabilizes public goods provision[*]

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

April 7, 2014

Abstract

We study the efficiency and stability properties of *meritocratic matching* in the context of group formation and public goods provision. This institutional mechanism is meritocratic in that it tends to assortatively match agents into groups according to their contributions. However, we assume that the correlated matching process is imperfect and probabilistic. The two extremes of our mechanism are the *voluntary contributions mechanism* (Isaac et al., 1985) with random group re-matching at the one end, and the *group-based mechanism* (Gunnthorsdottir et al., 2010a) at the other. The characteristics of meritocratic matching as a function of its degree of imperfection summarize as follows: (1) When matching is not sufficiently meritocratic, the only equilibrium state is universal free-riding. (2) Above a first threshold of minimum meritocracy, several Nash equilibria above free-riding emerge, but only the free-riding equilibrium is stochastically stable. (3) 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 new, more efficient equilibria with higher expected payoffs for all players. An important feature of the mechanism broadly speaking is that both groups of players, those that are incentivized by the mechanism to contribute as well as those that are not incentivized by the mechanism and continue to free-ride, will benefit from meritocratic matching.

JEL classifications: C73, D02, D03, D63

Keywords: Meritocracy, Voluntary contribution, Public goods, Stochastic stability, Evolutionary stability, Assortative matching

[*] The authors would like to thank Lukas Bischofberger for help with simulations, Bary Pradelski, Anna Gunnthorsdottir and Stefan Seifert for helpful comments on earlier drafts, Matthias Leiss for help with expressions (11)-(13), Peyton Young for help with framing of the questions, Luis Cabral for help with proposition 5, Ingela Alger and Jörgen Weibull for a helpful discussion, and finally the members of SOMS and DGBT at ETH Zurich, in particular Michael Mäs and Thomas Chadefaux, and the participants at "Norms, Actions, Games 2014" for helpful discussions. All remaining errors are ours. HHN acknowledges support by the European Commission through the ERC Advanced Investigator Grant 'Momentum' (Grant No. 324247).

[†] Corresponding author: Department of Social Sciences-SOMS, ETH Zürich, Clausiusstrasse 50-C3, 8092 Zürich, Switzerland. hnax@ethz.ch.

[‡] Department of Social Sciences-DBGT, ETH Zürich.

[§] Department of Social Sciences-SOMS, ETH Zürich.

1 Introduction

Meritocracy incentivizes effort and performance, thus promising efficiency gains. In environments such as education, job matching, or marriage markets, however, more meritocracy may lead to an exacerbation of inequality because the gap between over- and under-performers widens due to these added incentives (Young, 1958; Arrow et al., 2000; Greenwood et al., 2014). In this paper, we study the effects of meritocratic regimes that assortatively match players by their contributions to a public good. In particular, we apply meritocratic matching to games of public goods provision based on voluntary contributions, where non-meritocratic matching leads to free-riding and phenomena related to "tragedy of the commons" (Hardin, 1968; Ostrom et al., 1992). In this paper, we shall analyze the conditions under which sufficiently meritocratic matching can stabilize the provision of a public good. Moreover, we shall show that meritocratic matching in the context of public goods, in contrast to the environments previously studied, incurs little or no welfare costs of inequality because non-free-riding equilibria *ex-ante* payoff-dominate (Harsanyi and Selten, 1988) the free-riding equilibrium.

Specifically, we consider mechanistic variants of the public goods games introduced by Isaac et al. (1985) (see also Isaac and Walker 1988). Our mechanism differs substantially from the original set-up in several key aspects, but the common feature remains that individual players make voluntary contributions to a group account which then returns even shares of the group earnings back to the individual players in the group.[1] The unfortunate feature of the voluntary contributions game under the standard mechanism (Isaac et al., 1985) is that its unique equilibrium outcome is characterized by universal non-provision of the public good. We shall use the terminology *voluntary contributions mechanisms* (VCM) to refer to implementations of the contributions game by Isaac et al. (1985) where several separate groups are randomly matched from a wider population to produce several, separate, local public goods (e.g. Andreoni, 1988). Random re-matching à la Andreoni (1988) does not improve the game's pessimistic equilibrium predictions; the only Nash equilibrium remains to be global non-contribution. We shall take a different approach with respect to group configurations in this paper in that group matchings will not be random. Instead, a meritocratic regime will assortatively match groups based on players' decisions whether to contribute or not. Contributors will tend to be matched with contributors, and free-riders with free-riders. We shall call this contribution-assortative regime '*meritocratic matching*', abbreviated 'MERIT'. We study MERIT's efficiency and stability properties as a function of the protocol's randomness reflecting its meritocratic matching fidelity. MERIT nests what we defined as VCM above as the non-meritocratic limiting case. The other limiting case is the *group-based mechanism* (GBM) (Gunnthorsdottir et al., 2010a), which in our set-up represents the perfectly meritocratic case, and we shall elaborate on this case in more detail shortly.

But before we get into these details, we would like to provide more intuition for the basic flavour of the class of imperfect meritocratic matching mechanisms described by MERIT. While none of the following real-world institutions coincides one-to-one with MERIT, all of them share one of more of MERIT's key features. Entrance examinations to schools or universities, for example, assort individuals based on an imperfect measure of applicants' adequacies for different

[1] These voluntary contribution situations are related in spirit to the aforementioned common-pool resource problems (e.g. Ostrom et al., 1992; Ostrom, 1990), but one should not think of them as equivalent due to important psychological differences (Andreoni, 1995a).

streams of education. An important feature of this sorting mechanism is that the resulting differences in educational quality amongst those different schools are not only determined by the institutional design, but also largely 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 to get into a good school. We shall develop a formal model of such a matching process in the context of public goods games with *ex ante* homogeneous agents. Alternatively, imagine an assortative employment regime with team-based payments that reward employees for performance by matching them with similarly performant employees. Examples include trading desks in large investment banks, and again this 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, as with any real-world situation, all of these mechanisms are typically both noisy and not always fair.

So what is known about the effects of meritocratic matching in public goods situations? As mentioned before, standard VCM is non-meritocratic in terms of matching because players' contributions have no effect on their group memberships. Under VCM, numerous experimental and empirical studies show that contributions deteriorate towards the tragedy of the commons outcome without mechanistic additions (see Ledyard (1995) and Chaudhuri (2011) for reviews).[2] The recently proposed GBM (Gunnthorsdottir et al., 2010a), by contrast, is perfectly matching-meritocratic in that high contributors are guaranteed to join groups with aggregate contribution levels that are typically higher and definitely not lower than those of players with lower contributions. As a result, contributions under GBM gain additional group-matching advantages, and more efficient equilibria are enabled if the marginal rate of return is not too low (Gunnthorsdottir et al., 2010a). Several recent laboratory studies show that contributions are consistently stabilized by GBM in this case (Gunnthorsdottir et al., 2010a; Gunnthorsdottir and Thorsteinsson, 2010; Gunnthorsdottir et al., 2010b), and also by a pairwise demotion-promotion variant (Cabrera et al., 2013).

MERIT, the mechanism proposed here, bridges VCM and GBM. In the spirit of VCM and GBM, our mechanism also affects only group interactions, hence, there are no inherently mechanistic efficiency costs. There are also no payoff transfers via mechanistic additions such as taxes or subsidies, and no payoff can be used to pay for signals or punishment. Instead, players are assortatively matched based on their choices whether to contribute or not. Contributing and free-riding respectively imply the following new up- and downsides: contributing increases the chance of being matched into groups with high contributions; free-riding increases the risk of being matched in groups with low contributions. The size of these additional incentives depends on the level of meritocracy in matching. As a result, free-riding may seize to be a dominant strategy if sufficiently many players contribute, and if the system is sufficiently meritocratic. Indeed, MERIT enables equilibria above free-riding, which *ex ante* payoff-dominate the free-riding equilibrium.[3]

We model MERIT in such a way that the correlation of contribution decisions and group match-

[2] Other mechanisms that stabilize contributions include tax schemes (e.g. Tiebout, 1956; Buchanan, 1965; Groves and Ledyard, 1977), costly signalling (e.g. Gintis et al., 2001) and costly punishment (e.g. Ostrom et al., 1992; Fehr and Gaechter, 2000), all of which effectively change the payoff structure of the game.

[3] *Ex post*, there are subtle distributive issues which we shall address later in the paper.

ings is typically achieved imperfectly, e.g. due to monitoring issues, ill-defined measurements, or incomplete information. This generalization is important to understand how robust the effects of assortative matching are to imperfections from noise, errors, or misperceptions. Under imperfectly meritocratic matching, we shall therefore address the following questions: What are the conditions and, in particular, minimum levels of meritocratic matching fidelity necessary to generate equilibria that avoid tragedies of commons? And what are the efficiency, welfare and stability properties of these equilibria?

MERIT's meritocratic matching fidelity is described by parameter $\beta \in [0, 1]$ characterizing a class of mechanisms ranging from perfectly random to perfectly meritocratic. When $\beta = 0$, players are randomly matched, that is, chosen contributions have no effect on group memberships. This case corresponds to VCM, and free-riding is the unique Nash equilibrium. When $\beta = 1$, groups are formed according to an order of players' contributions that perfectly represents their magnitudes. This case corresponds to GBM (Gunnthorsdottir et al., 2010a), where near-efficient non-free-riding equilibria are shown to exist. Between the two extremes of no non-meritocratic matching and perfectly meritocratic matching, we investigate the cases where $\beta \in (0, 1)$.

The remainder of this paper is structured as follows. Next, we discuss the related literature, including the broad conceptual approach and the more specific models of public goods provision. In section 3, we develop a formal model of MERIT, analyze its equilibria, and detail their stability properties in an evolutionary setting. We conclude in section 4.

2 Related literature

2.1 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.[4] Until today, meritocratic institutions like the Chinese civil service examination are in place. Other modern examples include honorary circles, bonus wage schemes, etc.

The scope of this paper is limited to a class of public goods games, that is, 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

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

meritocratic incentives require welfare transfers, there is an inherent trade-off between efficiency and welfare. Previous research has identified precisely this feature as the main weakness of meritocratic regimes; i.e. that meritocracy increases heterogeneity of payoffs, and thus leads to inequality. Indeed, most of the modern political philosophy of meritocracy has focussed on this issue by analysis of situations with such an inherent efficiency-versus-equality trade-off (Young, 1958; Arrow et al., 2000). In the often-discussed context of education, for example, it is argued that merit-based reward first leads 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 meritocracy *per se*, but rather the result of interpreting what constitutes "merit" without distributional concerns. This distinguishes these previously studied meritocracy incentive structure from those present in non-constant-sum situations, which shall be the focus of this paper. Naturally, distributive costs may be substantial when meritocracy induces little efficiency gains or even none as in constant-sum environments. In strictly non-constant sum contexts such as public goods production which we shall consider here, however, our kind of meritocratic matching regime turns out to be a mechanism that is able to promote large efficiency gains at no or little distributive cost.

In this paper, we provide a stochastic model of meritocratic group matching in public goods situations based on players' contribution decisions. This distinguishes our model from an individual-based meritocracy model, where, for example, each individual player would receive a fixed bonus for every additional unit of contribution.[5] Instead, the role of meritocracy in our model is to match players into groups based on their contributions. The precision of this matching corresponds to the *level of meritocracy* of the system. One equilibrium of the game is characterized by uniform non-contribution. This state is maximally inefficient but perfectly equalitarian. In sufficiently meritocratic environments, other equilibria exist, where only a small fraction of players free-rides while the majority of players contributes. The new equilibria *ex ante* payoff-dominate the free-riding equilibrium. *Ex post*, however, players are categorized in three groups, the smallest of which will actually turn out to be worse-off. The first group are the free-riders. They are substantially better-off because expected contributions increase in all groups, so they will free-ride on higher contributions. Group two contains the majority of contributors who are also better-off, because they are matched in groups with high contributions. Group three contain the minority of contributors who are matched in groups with low contributions, and these 'unlucky' few will be slightly worse-off than in the free-riding equilibrium. Under severe *ex-post* equality concerns (e.g. Rawlsian), therefore, will meritocracy lead to welfare decreases. *Ex ante*, however, and for less-than-extreme levels of *ex-post* inequality aversion, will meritocracy be regarded as welfare-increasing.

[5] In such a model, individuals' decisions would be driven by assessments of marginal costs and benefits of contributing without any concern for group-formation effects.

2.2 Public goods, assortative matching and preference evolution

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 can lead to pro-social behavior in many situations (e.g. Hamilton 1964a,b; Nowak 2006; **?**). There are two recent papers studying the evolution of such preferences under different *assortative matching mechanisms*, that are in particular applied to public goods provision situations.[7] Alger and Weibull (2013) and Grund et al. (2013) 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*: in the sense of "homophily" in Alger and Weibull (2013), or "locality" in Grund et al. (2013). Such studies complement the analysis of biological mechanisms of kin selection in human interactions.

Both papers are motivated by arguments that do not presuppose the existence of a meritocratic institution, but rely on "higher-order" preference evolution driven by assortative matching. Given the payoff structure of the public goods game, sufficiently other-regarding preferences will incentivize voluntary contributions because players will seize to maximize only their own material payoffs. Grund et al. (2013) simulate agentic play of a voluntary contributions game on a lattice, and study the relative fitness of different embodied preferences as the players interact locally in neighborhoods, rather than with randomly drawn players from the global population. It is shown that local interactions facilitate the emergence and survival of agents with pro-social preferences (i.e., other-regarding '*homo socialis*' preferences). Alger and Weibull (2013) consider an alternative matching regime where individuals have a tendency to sort based on preference similarities (i.e. interaction homophily). Under this assumption, they find long-term survival of populations with *categorically imperative* ('*homo moralis*'; Kant 1785) utility concerns characterized by *homo hamiltoniensis* (Hamilton, 1964a,b) as a function of the degree of homophily.

Models like Alger and Weibull (2013) and Grund et al. (2013) show how non-selfish preferences may emerge. Non-selfish preferences are evolutionarily successful in these models because non-selfish populations evolutionarily outperform the *homo-oeconomicus* population. There are two main drivers for this. First, non-selfish agents interact mostly with each other, and thus attain payoffs that are superior to the *homo oeconomicus*. Second, even though the *homo oeconomicus* has higher payoffs than non-selfish players in his rare direct interactions with them, he is stuck most of the time with relatively low payoffs from interactions with other selfish agents. Negatively put, the models of Alger and Weibull (2013) and Grund et al. (2013), therefore, jointly imply that, if the interactions are sufficiently mixing (because the game is globally too integrated or because there is too little homophily), then cooperation will not emerge because the *homo-oeconomicus* population will always outperform any non-selfish population.

The interesting link with our model is that the phenomena in terms of players' decisions in Alger and Weibull (2013) and Grund et al. (2013) could be re-interpreted *as if* players with

[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.

[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.

homo-oeconomicus preferences interacted in a meritocratic world without any lattice structure, homophily or *non-homo-oeconomicus* agents. If one interpreted only the correlation between players' chosen actions and resulting interactions in these models, it is *as if* the two models generate meritocratic matching: essentially, contributors and free-riders have a tendency to be rematched with players doing likewise. However, in these models this is not driven because players' actions are evaluated by some exogenous institution (as in our model), but due to the game's lattice structure or agents' homophilic tendencies. The phenomenon of *as-if* meritocratic matching in the long-run stable states in these models is then a consequence of emergent *non-homo-oeconomicus* preferences and the mixing constraints. This implies, however, that if a player in the models of Alger and Weibull (2013) and Grund et al. (2013) was to mutate from a non-selfish player into *homo oeconomicus* again, he would always choose to free-ride.

That fact that *homo oeconomicus* rationally chooses to contribute is the key feature of our model and the key difference with the models of Alger and Weibull (2013) and Grund et al. (2013). The common feature lies in the assortativity of matching driving the results. 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. MERIT 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, or hope to be 'recognized' for contributing, or fear to be 'stigmatized' for free-riding (e.g. Andreoni and Petrie, 2004; Samek and Sheremeta, 2014), that is, our mechanism functions 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 Gaechter, 2000; Fehr and Camerer, 2007; Fehr and Schmidt, 1999; Fehr and Gaechter, 2002) to our kind of meritocratic matching environment.

2.3 Contributions mechanisms in experimental economics

Our model contributes to the theoretical underpinnings of the literature on public goods games with voluntary contributions 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) and Isaac and Walker (1988). Within that literature, the research on group formation is most closely related to this paper. In the initial models, the groups within which the public good was provided remained fixed (Isaac et al., 1985). To disentangle learning drom reputation effects in repeated games, this mechanism was extended to random group re-matching (e.g. Andreoni, 1988).[9] An important avenue has been to model how groups may form endogenously. Several such mechanisms have been proposed. Cinyabuguma

[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).

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

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 the non-random group dynamics stabilize higher contributions.

Most closely related to our model is the *group-based mechanism* (GBM; Gunnthorsdottir et al. 2010a). In fact, in MERIT we propose a mechanism that bridges the *voluntary contributions mechanism* (VCM; e.g. Andreoni 1988) and GBM. These two ends of our model spectrum we shall now detail. Random group re-matching as in VCM is the baseline model, defining the *no-meritocracy* end of the continuum of MERIT. VCM proceeds as follows. First, n players make voluntary contributions. Second, they are randomly sorted into n/s groups each composed of s players. Importantly, players' chosen strategies 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 (MPCR), which aggregates the contributions in each group, multiples the sum by a coefficient greater than one, and then evenly divides the product among the group members. For MPCR in between $1/s$ (a fraction proportional to the group size) and one, total welfare is highest if everybody makes the maximal contribution, but, given contributions of others, each player maximizes his own payoff by contributing nothing; as a result, free-riding is the unique dominant strategy.

Under GBM (Gunnthorsdottir et al., 2010a), players are ranked in order of magnitude of contributions and subsequently matched into groups based on that order. GBM represents *perfect meritocracy* in our model. It proceeds as follows. First, players make voluntary contributions. Second, players are ranked such that no player who contributes less than another is ranked higher than him (ties are randomly broken).[10] Third, groups of fixed size s form by rank; the s top contributors form group one, the next s contributors form group two, etc. Finally, payoffs realize dependent on the MPCR $\in (1/s, 1)$. The free-riding Nash equilibrium (NE) continues to exist, but other equilibria may emerge depending on population, group size and MPCR. The focus of the paper by Gunnthorsdottir et al. (2010a) is the near-efficient NE in which a large majority of players contributes everything and only a small fraction of players free-rides.

We propose MERIT, and thus fill the space between VCM and GBM with a *stochastically meritocratic matching mechanism.* Typically, MERIT is neither completely random nor perfectly meritocratic. Players are ranked by their contributions such that a contributor is more likely to be ranked higher than a free-rider, but with some positive probability he is ranked lower. If we consider β as the level of meritocracy in the system, then at one extreme, when $\beta = 0$ (no meritocracy), the mechanism is VCM, and, at the other extreme, when $\beta = 1$ (perfect meritocracy), the mechanism is GBM.

We shall now proceed to formalize the set-up of our model.

[10] Note that GBM, which corresponds to 'perfect' meritocracy in our model itself already bears an element of probabilistic chance due to this element of random tie-breaking, which will mean that 'unlucky' contributors may end up in groups with players who contribute less.

3 Meritocratic matching in voluntary contributions games

3.1 A voluntary contributions game with meritocratic matching

Suppose population $N = \{1, 2, ..., n\}$ plays the following meritocratic matching contributions 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 $c = \{c_i\}_{i \in N}$.

Step 2. Ordering as a function of contributions

An authority imperfectly observes the contribution vector c and imperfectly ranks players accordingly. The measure of ranking precision is given by parameter $\beta \in [0, 1]$. The characteristics of the regimes summarize as follows: *(i) without meritocracy* ($\beta = 0$), all rankings are equally likely, and all players have the same expected rank; *(ii)* in *perfect meritocracy* ($\beta = 1$), all rankings are perfect, and all contributors 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 higher contributors have a higher expected rank than free-riders.

We shall now proceed to formalize this.

Rank orderings. 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).

MERIT. 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; an interesting function to analyze is $\overline{k}_i^\beta(c_i = 0) - \overline{k}_i^\beta(c_i = 1)$, 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 characterize MERIT:

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

(ii) perfect 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 $c_j = 1$

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

$$\partial \left(\overline{k}_i^{\beta}(c_i = 0) - \overline{k}_i^{\beta}(c_i = 1) \right) / \partial \beta > 0, \tag{1}$$

$$\overline{k}_i^{\beta}(c_i = 1) < \overline{k}_i^{\beta}(c_i = 0). \tag{2}$$

That is, a player's expected rank difference is always positive and increasing in β. There are many functional assumptions that satisfy these requirements, one of which is the following:

MERIT 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

$$p_i(k) = \frac{e^{l_i}}{\sum_{j \in N \setminus S} e^{l_j}}. \tag{3}$$

Other interpretations of $\beta \in [0, 1]$ are *(i)* β represents the fraction of every contributed unit to enter GBM and $1 - \beta$ to enter VCM, or *(ii)* $(1 - \beta)/\beta$ represents some normally distributed noise δ^2 added to the contribution vector c after which GBM is applied to $x \sim N(c, \delta^2)$.

Step 3. Grouping as a function of orderings

Finally, groups form based on the ranking and payoffs realize based on the contributions made in each group.

Groupings. 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* (MPCR) r/s, i receives

$$\phi_i(c_i|c_{-i}, \rho) = \underbrace{(1 - c_i)}_{\text{remainder from budget}} + \underbrace{\sum_{j \in S} (r/s) c_j}_{\text{return from the public good}}. \tag{4}$$

It is standard to assume that $r/s \in (1/s, 1)$, in which case contributing is socially beneficial. But note that, in the non-meritocratic-matching case, individual incentives lead to tragedy of the commons (details are provided in the analysis of the Nash equilibria in the next section).

3.2 Nash equilibria

From expression (4), the expected payoff of contributing c_i given c_{-i} for any i is

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

where $S_i^{\pi} \in \rho$ is subgroup to which player i belongs. Note that term (iii), the expected return from others' contributions, is a function of one's own contribution due to meritocratic matching. We shall use expression (5) to make the following equilibrium observations.

Proposition 1. *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 NE*** *(FRNE) such that all players free-ride.*

Proof. The proof of Proposition 1 follows from the fact that, given c_{-i} such that $c_j = 0$ for all $j \neq i$, we have:

$$1 = \mathbf{E}\left[\phi_i(0|c_{-i})\right] > \mathbf{E}\left[\phi_i(1|c_{-i})\right] = r/s.$$

□

The proof follows from the fact that it is never a best response to be the only contributor. Note that if, for all i, given any c_{-i} and β, $\mathbf{E}\left[\phi_i(0)|c_{-i}\right] > \mathbf{E}\left[\phi_i(1)|c_{-i}\right]$, then we have a situation where free-riding is the dominant strategy for any level of meritocracy.

Write 1^m for "m players contribute, all others free-ride", and 1^m_{-i} for the same statement excluding player i. Write $mpcr_1$ for the marginal per capita rate of return $r/s = \frac{n-s+1}{ns-s^2+1}$.

Proposition 2. *Given population size $n > s$, group size $s > 1$ and rate of return r such that $r/s \in (mpcr_1, 1)$, there exists $\underline{\beta} \in (0, 1)$ above which there is a* ***pure-strategy NE*** *(PSNE), where $m > 0$ agents contribute and the remaining $n - m$ agents free-ride.*

Proof. The following two conditions must hold for Proposition 2 to be true:

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

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

The proof for the existence of an equilibrium with $m > 0$ when $\beta = 1$ follows from Theorem 1 in Gunnthorsdottir et al. (2010a), in which case both equations (6) and (7) are strictly satisfied if $r/s > mpcr_1$.

The fixed point argument behind Theorem 1 in Gunnthorsdottir et al. (2010a) becomes clear by inspection of terms *(ii)* and *(iii)* in expression (5): 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 $(n-1)$ (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 NE for $\beta = 1$ with $0 < m < n$ and $r/s \in (mpcr_1, 1)$ in which case equations (6) and (7) 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:

$$\partial \mathbf{E}\left[\phi_i(1|1^m_{-i})\right]/\partial\beta > 0 \tag{8}$$

$$\partial \mathbf{E}\left[\phi_i(0|1^m_{-i})\right]/\partial\beta < 0 \tag{9}$$

When $\beta = 0$, we know that $\phi_i(1|1^m_{-i}) = r/s < \phi_i(0|1^m_{-i}) = 1$ for any m. However, by existence of the equilibrium with $m > 0$ contributors when $\beta = 1$, provided that $r/s > mpcr_1$ is satisfied, there must exist some maximum value of $\underline{\beta} \in (0,1)$, at which either equation (6) or equation (7) first binds due to continuity of expressions (8) and (9). That level is the bound above which the PSNE with $m > 0$ exists. □

Note that, for a finite population of size n, a group size s larger than one implies that $r/s > 1/s$ for Proposition 2 to be true, but as $n \to \infty$, the range of r/s converges to $(1/s, 1)$.

A special case of a PSNE is the **near-efficient pure-strategy NE (NEPSNE)** (see Gunnthorsdottir et al., 2010a): for any $\beta > \underline{\beta}$, NEPSNE is the PSNE such that m is chosen to be the largest value given n, s, r for which equations (6) and (7) hold.

Let $p_i \in [0,1]$ be a mixed strategy with which player i plays contributing ($c_i = 1$) while playing free-riding with $(1-p_i)$ ($c_i = 0$). Write $p = \{p_i\}_{i\in N}$ for a vector of mixed strategies. Write 1_p for “all $j \neq i$ play p”, and 1^p_{-i} for the same statement excluding player i.

Proposition 3. *Given population size $n > s$ and group size $s > 1$, there exists a rate of return r such that $r/s \in [mpcr_1, 1)$ beyond which there exists $\underline{\beta} \in (0,1)$ such that there always exist two mixed strategy profiles, where every agent places weight $p > 0$ on contributing and $1-p$ on free-riding, that constitute* ***symmetric mixed-strategy NE*** *(SMSNE), one with a high $\overline{p}$ (the* near-efficient *SMSNE) and one with a low $\underline{p}$ (the* less-efficient *SMSNE).*

Proof. The SMSNE exists if there exists a $p \in (0,1)$ such that

$$\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] = \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]. \tag{10}$$

In that case, player i has a best response playing $p_i = p$ which would be a Nash equilibrium. Proposition 2 implies that, if $r/s > mpcr_1$, equations (6) and (7) are strictly satisfied when $\beta = 1$ for m contributors corresponding to NEPSNE. Indeed, expressions (6) and (7) imply a lower bound, l, and an upper bound, u, for the number of free-riders, $(n-m)$, given by (see Gunnthorsdottir et al., 2010a)

$$l = \frac{n - nr/s}{1 - r/s + nr/s - r}, \qquad u = 1 + \frac{n - nr/s}{1 - r/s + nr/s - r}. \tag{11}$$

Part 1. First, we will show, for the case when $\beta = 1$, that there is at least one SMSNE when $r/s \to 1$, possibly none when $r/s = mpcr_1$, and that there is a continuity in r/s such that there is some intermediate value of $r/s \in [mpcr_1, 1)$ above which at least one SMSNE exists but not below.

First, because $\partial \mathbf{E}\left[\phi_i(c_i|1^p)\right]/\partial p > 0$ for all c_i, there exists a $p \in (\frac{m-1}{n}, \frac{m+1}{n})$ such that expression (10) holds if $r/s \to 1$. This is the standard symmetric mixed-strategy Nash equilibrium in a symmetric two-action n-person game when the only pure-strategy equilibria are asymmetric (see the proof of Theorem 1 in Cabral 1988). In this case, the presence of the FRNE makes no difference because the incentive to free-ride vanishes as $r/s \to 1$.

Second, if $r/s = mpcr_1$, one or both of the equations, (6) or (7), bind. Hence, unless expression (10) holds exactly at $p = m/n$ (which is a limiting case that we will address in proposition (5)), there may not exist any p such that expression (10) 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 FRNE while universal contributing is not a Nash equilibrium. In this case, the incentive to free-ride is too large.

Third, $\partial \mathbf{E}\left[\phi_i(c_i|1^p_{-i})\right]/\partial r$ is a linear constant > 0 for both $c_i = 0$ and $c_i = 1$. At and above some intermediate value of r/s, therefore, there exists a $p \in (0, 1)$ such that, if played in a SMSNE, the incentive to free-ride is mitigated sufficiently to establish equation 10.

Finally, for any $p > 0$ constituting a SMSNE 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 2 applies to ensure the existence of some $\underline{\beta} \in (0, 1)$ above which the SMSNE continues to exist when $r/s > mpcr_2$: because, at $\beta = 1$, equations (6) and (7) are strictly satisfied and $\mathbf{E}\left[\phi_i(0|1_p)\right] = \mathbf{E}\left[\phi_i(1|1_p)\right] > 1$, there exist some $\beta < 1$ and $p' < p$ satisfying equation (10 while still satisfying $\mathbf{E}\left[\phi_i(0|1_p)\right] = \mathbf{E}\left[\phi_i(1|1_p)\right] > 1$.

Part 2. If $r/s > mpcr_2$ and $\beta > \underline{\beta}$, existence of two equilibria with $\overline{p} > \underline{p} > 0$ is shown by analysis of the comparative statics of equation (10).

First note that, for any $r/s > mpcr_2$ 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 (10) to hold for two different values of β above $\underline{\beta}$. 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 (10), and obtain

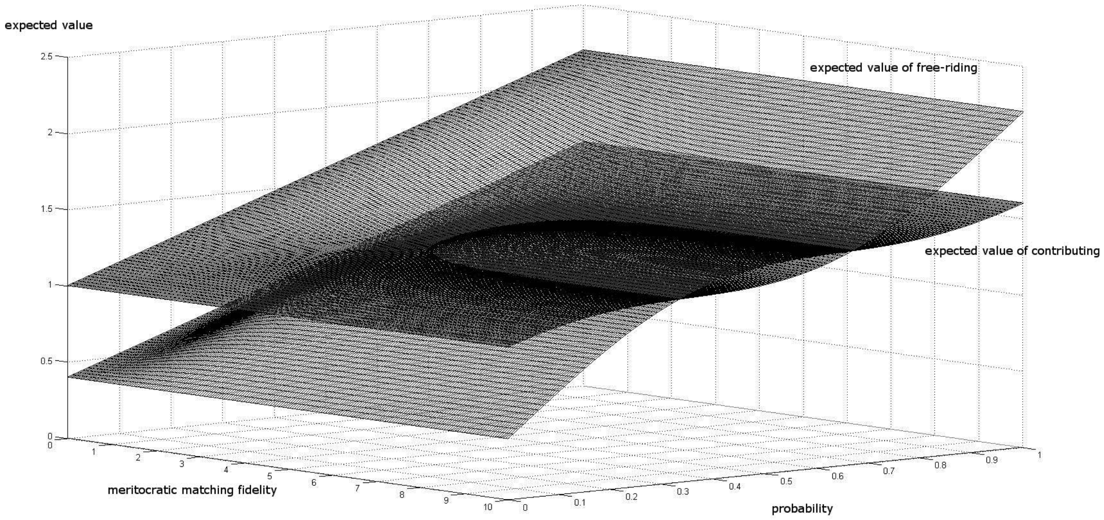
$$\partial p/\partial \beta = \frac{\partial \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]/\partial \beta - \partial \mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial \beta}{\partial \mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial p - \partial \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]/\partial p}. \tag{12}$$

Expression 12 is negative if the denominator is negative, because the numerator is always positive.

Claim 4. *The denominator of equation* (12) *is negative when p is low, and positive when p is high.*

Write $\overline{w}^i_{c_i}$ and $\underline{w}^i_{c_i}$ respectively for the probabilities with which agent i is matched in an above- or below-average group when playing c_i. Write $\mathbf{E}\left[\overline{\phi}_i(c_i|1_p)\right]$ and $\mathbf{E}\left[\underline{\phi}_i(c_i|1^p_{-i})\right]$ for the corresponding expected payoffs.

Figure 1: Contributing versus free-riding ($n = 16$, $s = 4$, $r = 1.6$).



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 MERIT via logit (equation* (3)*). The two planes intersect at the bifurcating SMSNE-values of* $\overline{p}$ *and* $\underline{p}$ *(see Proposition* (3)*). Notice that the curves are linear when the meritocratic matching fidelity is zero but turn S-shaped for larger values, thus intersecting at* $\overline{p}$ *and* $\underline{p}$*.*

Recall that, for $\beta > 0$ and $1_p \in (0, 1)$, expression (1) holds, where $\widehat{k}$ is compatible with a perfect ordering $\widehat{\pi}$, and $\widetilde{k}$ is any rank compatible with a mixed ordering $\widetilde{\pi}$. When $1_p = 0$ or 1, f^{ij} is independent of c_i, and in particular $\overline{w}^i_{c_i} = \underline{w}^i_{c_i}$ for any c_i.

Hence, we can rewrite $\partial \mathbf{E}\left[\phi_i(0|1^p_{-i})\right]/\partial p$ in the denominator of equation (12) as

$$\partial \overline{w}^i_0/\partial p \cdot \mathbf{E}\left[\overline{\phi}_i(0|1^p_{-i})\right] + \partial \underline{w}^i_0/\partial p \cdot \mathbf{E}\left[\underline{\phi}_i(0|1^p_{-i})\right] \tag{13}$$

and $\partial \mathbf{E}\left[\phi_i(1|1_p)\right]/\partial p$ as

$$\partial \overline{w}^i_1/\partial p \cdot \mathbf{E}\left[\overline{\phi}_i(1|1^p_{-i})\right] + \partial \underline{w}^i_1/\partial p \cdot \mathbf{E}\left[\underline{\phi}_i(1|1^p_{-i})\right]. \tag{14}$$

It follows from continuity that the denominator of equation (12) is negative when p is low, and positive when p is high. □

Proposition 5. *Given group size $s > 1$, then, if $\beta = 1$, as $n \to \infty$ (i.) $1_k/n$ of NEPSNE and $\overline{p}$ of the near-efficient SMSNE converge, and (ii.) the range of r/s for which these equilibria exist converges to $(1/s, 1)$.*

Proof. Suppose r/s is such that both SMSNE and NEPSNE exist. Let 1_k describe NEPSNE. Recall that expressions under (11) summarize the lower bound, l, and upper bound, u, for the number of free-riders $n - 1_k$ under NEPSNE. Taking $\lim_{n\to\infty}$ for the equations under (11) imply a limit lower bound of $\frac{1}{1+n\frac{r/s-r/n}{1-r/s}}$, and a limit upper bound of the expected proportion of free-riders of $\frac{1}{n} + \frac{1}{1+n\frac{r/s-r/n}{1-r/s}}$, and 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 $r/s \to 1$.) We know that, if there is one more free-rider than given by the upper bound, then equation 7 is violated. Similarly, if there is one fewer free-rider than given by the lower bound, then equation 6 is violated.

Let 1_p describe the near-efficient SMSNE. Recall that $\mathbf{E}\left[\phi_i(0|1^p_{-i})\right] = \mathbf{E}\left[\phi_i(1|1^p_{-i})\right]$ (expression 10) must hold for any player i given all $j \neq i$ play p, where $\mathbf{E}\left[\phi_i(c_i|1^p_{-i})\right] = \mathbf{E}\left[\phi_i(c_i|1^b)\right]$ where 1^b is the proportion of other players actually contributing (playing one) which is distributed according to a Binomial with mean $\mathbf{E}[b] = p$ and variance $\mathbf{V}[b] = \frac{p(1-p)}{n-1}$. As $n \to \infty$, by law of large numbers, we can use the same bounds obtained for the NEPSNE to bound $p \in [(n-u)/n, (n-l)/n]$, which converge to the unique p at which expression 10 actually holds.[11]

Suppose all players contribute with probability p corresponding to the near-efficient SMSNE limit value. Then, $\lim_{n\to\infty} \mathbf{V}[b] = \lim_{n\to\infty} \frac{p(1-p)}{n} = 0$ for the actual proportion of contributors b. Hence, the limit for the range over r/s necessary to ensure existence converges to that of the NEPSNE, which following the proof of Theorem 1 in Gunnthorsdottir et al. (2010a) is $(1/s, 1)$. □

Remark 6. *In light of the limit behavior, it is easy to verify that* $\partial mpcr_2/\partial n < 0$ *and* $\partial mpcr_2/\partial s > 0$*, .*

Summary of unilateral best replies. The results from Propositions 1 to 5 are summarized by the following four observations for the case when $r/s > mpcr_2$:

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 $1_m \geq 0$

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 $p \in (\underline{p}, \overline{p})$

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 $\overline{p}$

Figure 1 illustrates the *expected values* of $\phi_i(0|1^p_{-i})$ and $\phi_i(1|1^p_{-i})$ as functions of *probability* p with which players contribute, and *meritocratic matching fidelity* β under MERIT via logit (equation (3)) for $n = 16$, $s = 4$ and $r = 1.6$.[12]

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

[12] Parameter values used in related studies of VCM and GBM are $n = 12$, 16 or 20, $s = 2$ or 4, and $r = 1.6$ or 2 (see Ledyard 1995; Chaudhuri 2011 for the VCM, and Gunnthorsdottir et al. 2010a; Gunnthorsdottir and Thorsteinsson 2010; Gunnthorsdottir et al. 2010b for the GBM). Our numerical implementations and illustrations are done with $n = 16$, $s = 4$, $r = 1.6$ using MERIT via logit (equation (3)).

3.3 Efficiency and welfare

Table 1: Stem-and-leaf plot for FRNE and NEPSNE ($n = 16$, $s = 4$, $r = 1.6$, $\beta = 1$).

	NEPSNE	payoff	FRNE	
	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 with their contributions. At the bottom, the efficiencies of the two outcomes are calculated. Note that NEPSNE is more efficien, whereas FRNE is more equitable.

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

Efficiency. $\sum_{i \in N} \phi_i$ is the *efficiency* of outcome (ρ, ϕ).

Table 1 summarizes FRNE and NEPSNE when $\beta = 1$ for the economy with $n = 16$, $s = 4$ and $r = 1.6$. Table 2 lists the efficiencies of the different NE in general.

Table 2: Efficiency of equilibria $r/s \in (mpcr_2, 1)$.

equilibrium	existence condition	efficiency
FRNE	$\forall \beta$	$= n$
PSNE with $m > 0$ contributors	if $\beta > \underline{\beta}$	$= (n - m) + m \cdot r > n$
SMSNE with $p > 0$ for contributing	if $\beta > \underline{\beta}$	$\in [n, nr]^*$

*: $\mathbf{E}[\sum_{i \in N} \phi_i] = (1 - p) \cdot n + p \cdot n \cdot r > n$

Social welfare. Given outcome (ρ, ϕ), let $W(\phi)$ be a *social welfare function* (SWF) measuring the welfare of π.

One particular SWF is **Utilitarian** (Bentham, 1907):

$$W_U(\phi) = \frac{1}{n} \sum_{i \in N} \phi_i \tag{15}$$

Obviously, Utilitarianism maximizes efficiency favouring equilibria with high levels of contributions, because there are no concerns for distributional inequality under the Benthiam SWF criterion. Hence, Utilitarianism would prefer NEPSNE to any other equilibrium.

The other extreme is the **Rawlsian** SWF (Rawls, 1971):

$$W_R(\phi) = \min(\phi_i) \tag{16}$$

Rawlsianism associates welfare with the utility of the person who is *ex post* worst-off. In our game, the Rawlsian-optimal equilibrium is FRNE with perfect equality of payoffs (equal to one for every player). This is because, in every other candidate SMSNE or NEPSNE with positive contribution levels, every contributor, with strictly positive probability, receives a payoff of less than one. **Harsanyi**'s SWF (Harsanyi, 1953), on the other hand, would prefer NEPSNE and SMSNE to FRNE because every contributor and every free-rider is in expectation (i.e. *ex ante*) better-off in NEPSNE than in any other equilibrium.[13] Which equilibrium is preferable in terms of social welfare for any SWF depends on the 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).[14]

The **Cobb-Douglas** SWF nests both the Utilitarian and Rawlsian SWFs:

$$W_e(\phi) = \frac{1}{n(1-e)} \sum_{i \in N} \phi_i^{1-e} \tag{17}$$

where e represents the parameter of *inequality aversion* of the SWF with $e \in [0, \infty)$.[15] When $e = 0$, expression (17) reduces to expression (15), and when $e \to \infty$, expression (17) is approximated by function (16). In our game, the value of e determines whether a move from FRNE to one of the other equilibria would be *ex-post* welfare-enhancing or not. For the economy illustrated in Table 1 (with $n = 16$, $s = 4$ and $r = 1.6$), an *ex-post* Cobb-Douglas SWF comparison with $e < 10.3$ prefers NEPSNE to FRNE, while any Cobb-Douglas SWF with $e \geq 10.3$ prefers FRNE to NEPSNE. With inequity aversion of $e = 10.3$, the social planner requires efficiency gains of more than twice the amount lost by any player. *Ex ante*, of course, any Cobb-Douglas SWF prefers NEPSNE because of *ex-ante* payoff dominance.

3.4 Stability

Relative *ex-ante payoff dominance* and *risk dominance* (the size of the basins of attractions) ultimately determine the stability of the different equilibria. In this section, we shall analyze the stability properties of states as in *evolutionary stability* (Maynard Smith and Price, 1973) under replicator dynamics (Taylor and Jonker, 1978; Helbing, 1996) and *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 the policy choice. Depending on which welfare function is pursued, MERIT will stabilize different equilibria, and we want to understand how much meritocracy in matching is necessary and sufficient for stabilizing the different equilibria.

[13] Harsanyi's SWF is $W_H(\phi) = \frac{1}{n} \sum_{i \in N} \mathbf{E}[\phi_i]$.

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

[15] See, for example, Binmore (2005) for a discussion of the Rawlsian and Harsanyi 'original position' approach and the Cobb-Douglas generalization.

We shall begin by defining the following dynamic game played by myopic agents. 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 the proportion of p players contributing and the remaining $1 - p$ free-riding. Let Ω be the state space.

3.4.1 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):

$$\partial p/\partial t = (1-p)p\left(\mathbf{E}\left[\phi_i(1|1_p)\right] - \mathbf{E}\left[\phi_i(0|1_p)\right]\right) \tag{18}$$

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 $p > \bar{p}$.

Lemma 7. *Given population size $n > s$, group size $s > 1$ and rate of return r such that $r/s \in (mpcr_2, 1)$, there exists a $\underline{\beta} > 0$ below which the only ESS is FRNE. When $\beta > \underline{\beta}$, in addition, the population proportions given by the near-efficient SMSNE are also ESS.*

Proof. The proof of Lemma 6 and the cut-off structure of the ESS as given by Proposition 2 follow from Observations A-C: Observation A implies that the the only ESS when $\beta < \underline{\beta}$ is given by FRNE. Observations B1 implies that FRNE is also ESS when $\beta \geq \underline{\beta}$. Observation B1, B2 and C, jointly, imply that a population playing according to the contribution proportions given by the near-efficient SMSNE is also ESS. □

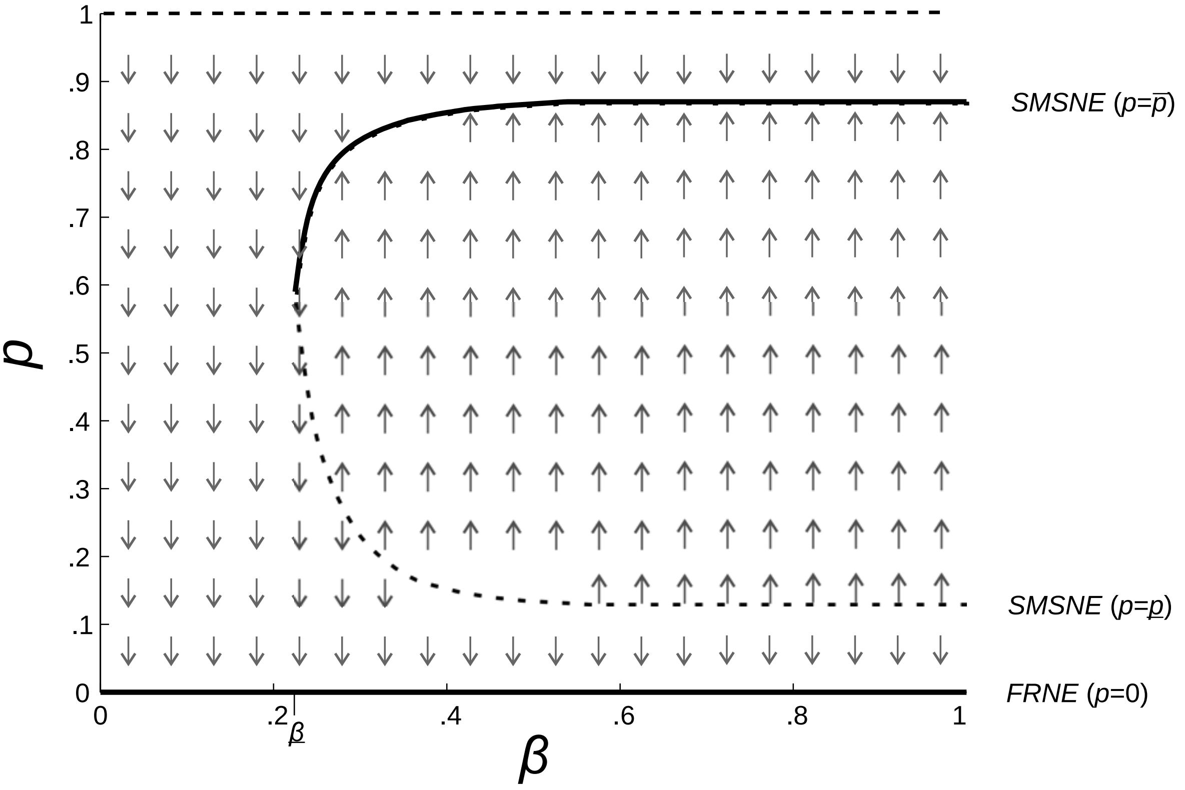
Remark 8. *As replicator dynamics increase n, recall that the bound on r/s converges to* $(1/s, 1)$ *(proposition* 5*). Hence, proposition* (7) *is a general observation about the near-efficient SMSNE for any rate of return.*

Figure 2 illustrates the implied replicator dynamic phase transitions for proportions of players contributing as a function of β under MERIT via logit (equation (3)) for $s = 4$ and $r = 1.6$ starting with $n = 16$. 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 'drop down' is sufficient to drop out of the high equilibrium into the free-riding equilibrium.

3.4.2 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, thus starting a new discrete time step t. When individual i is activated (the uniqueness of which is guaranteed by the independence of the Poisson clocks), all agents $j \neq i$

Figure 2: Evolutionary stability ($n = 16$, $s = 4$, $r = 1.6$).



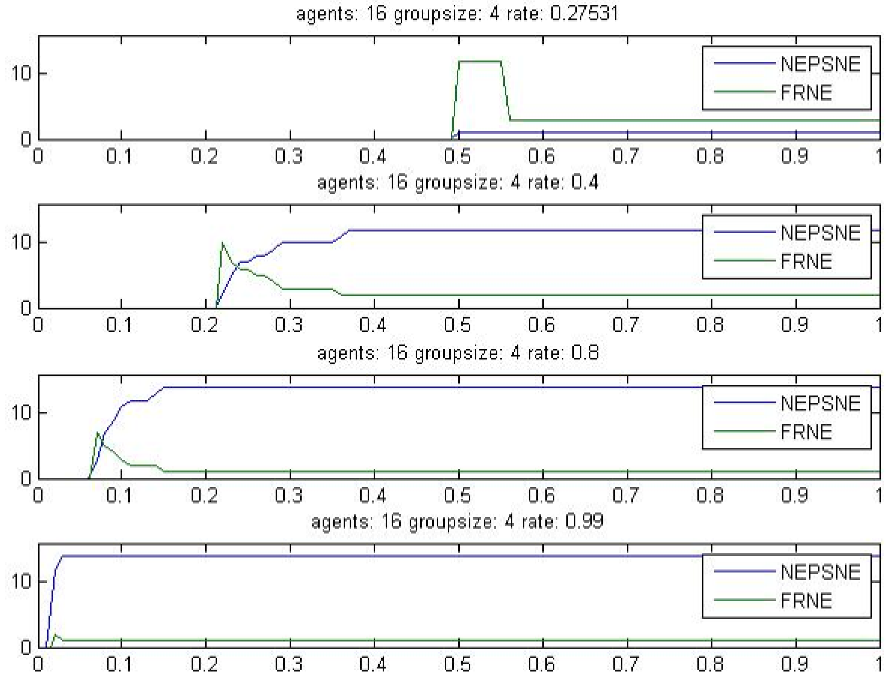
For $\beta < \underline{\beta}$, and for $\beta > \underline{\beta}$ when p is in excess of the near-efficient SMSNE ($\overline{p}$) but short of the less-efficient SMSNE ($\overline{p}$), $\partial p/\partial t < 0$ (replicator tendency is down). For $\beta > \underline{\beta}$ when $\overline{p} > p > \underline{p}$, $\partial p/\partial t > 0$ (replicator tendency is up). Depending on the location along the bifurcation, the evolutionarily stable states are therefore when either $p = 0$ (FRNE) and when p is set according to the near-efficient SMSNE ($\overline{p}$).

continue playing their previous strategy ($c_j^t = c_j^{t-1}$), while i plays 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.

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, the absorbing states of the unperturbed process (when $\epsilon = 0$) are the aforementioned FRNE and PSNE for any given level of β as identified in section 3.2. Now, we analyze the stability of the absorbing states based on equilibria's "critical mass" necessary to destabilize them:

Critical mass. The *critical mass* $\mathcal{M}_p^\beta \in [0, n-1]$ necessary to destabilize state p given β is defined as the minimum number of errors needed from any set of players $S \subset N$ such that playing the current strategy for at least one player in $N \setminus S$ is no longer a best reply.

Figure 3: Stochastic stability ($n = 16$, $s = 4$).



The two step functions in each panel illustrate the critical mass necessary to destabilize FRNE and NEPSNE in four different economies as a function of β. The economies have population size $n = 16$ and group size $s = 4$. Four rates of return r are chosen such that the MPCR is $r/s = 0.28$, 0.4, 0.8 or 0.99. For $\beta \to 0$, FRNE is the unique (stochastically stable) equilibrium for all r/s. In the first panel ($r/s = 0.28$), FRNE is the unique stochastically stable equilibrium for all β. In the other panels, NEPSNE becomes stochastically stable above some intermediate value of β. In the second and third panel, NEPNSE exists above some value of β and becomes stochastically stable above another. In the fourth panel, NEPSNE is stochastically stable right from the β above which it exists.

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 only equilibrium which is free-riding ($p = 0$); $\mathcal{M}_0^\beta = \emptyset$. When $\beta = \underline{\beta}$, the PSNE has a critical mass of $\mathcal{M}_{\underline{p}}^\beta = 1$. When $\beta > \underline{\beta}$, for any NE that is not either FRNE or NEPSNE (with $p = \underline{p}$), 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). For $\beta > \underline{\beta}$, $\partial\mathcal{M}_0^\beta/\partial\beta < 0$ and $\partial\mathcal{M}_{\overline{p}}^\beta/\partial\beta > 0$.

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

Lemma 9. *The unique stochastically stable state is either NEPSNE with $p = \overline{p}$ if $\mathcal{M}_0^\beta < \mathcal{M}_{\overline{p}}^\beta$ or FRNE with $p = 0$ when $\mathcal{M}_0^\beta > \mathcal{M}_{\overline{p}}^\beta$. When $\mathcal{M}_0^\beta = \mathcal{M}_{\overline{p}}^\beta$, both are stochastically stable.*

Proof. This is an 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. $\square$

Figure 3 illustrates the critical masses to destabilize FRNE and NEPSNE respectively as a function of β for four different economies. The economies vary in their rates or return. Whichever curve lies higher indicates the stochastically stable state; when both curves lie on the x-axis, FRNE is the unique (stochastically stable) equilibrium, in that case NEPSNE does not exist. The two lines meeting indicates where both states are stochastically stable. Note that the level of β necessary to destabilize FRNE is decreasing in r/s, hence decreasing in r but increasing in s.

4 Conclusion

Many real-world institutions have a meritocratic element (such as admission to university admission or civil service), including those that aim to incentivize contributions to a public good (such as sports or group efforts). In public goods games, a player's contribution has positive externality effects on others. Under standard mechanisms, and in particular when interacting players are homogeneously and randomly mixed, free-riding is a dominant strategy, and non-provision of the public good results. In this paper, we propose a mechanism that tends to assortatively match players with similar contributions together. That way, players who contribute more or less also tend to benefit more or less from others. This kind of meritocratic matching constitutes a competitive principle that overcomes the dilemma of non-provision resulting from random interactions in this class of games. Because measures and assessments of merit and implementations of meritocracy are often imperfect in reality, this paper analyzes which levels of meritocratic matching fidelity are necessary and sufficient to induce and stabilize equilibria with high contributions. The corresponding robustness depends on population size, group size, and rate of return. Moreover, when meritocratic matching leads to equilibria with positive contributions, it turns out that the resulting distributional costs are small compared to the efficiency gains. The reason for this is that the mechanism is group-based, not individual-based, and therefore not only benefits contributors but also generates externality benefits for free-riders, that is, for those players who are not incentivized to contribute by the mechanism. This reflects the fact that meritocratic matching, for the majority of players, mitigates the free-riding dilemma by incentivizing contributions, but, because such a mechanism does not change the basic payoff structure of the game, does not completely overcome this inherent incentive for all players.

It is an avenue for future research to study the effects of meritocratic matching in different and more general classes of games, and to address the co-evolution of meritocratic institutions with population geography, homophily and preferences. Our next step is to validate the model's imperfect meritocracy predictions in the economic laboratory. The *status quo* of experimental public goods studies can be described as follows. Without meritocracy, contributions deteriorate towards the free-riding NE in the standard setting without mechanistic additions such as punishment, signalling or taxes (see Ledyard, 1995; Chaudhuri, 2011). What we have described as 'perfect' meritocracy, on the other hand, consistently stabilizes near-efficient contributions (Gunnthorsdottir et al., 2010a; Gunnthorsdottir and Thorsteinsson, 2010; Gunnthorsdottir et al., 2010b).

References

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.

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.

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

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

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

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

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.

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

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

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

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, 1995a.

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

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.

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

C. Tiebout. A pure theory of local expenditures. *The Journal of Political Economy*, 64:416–24, 1956.

J. M. Buchanan. An economic theory of clubs. *Economica*, 32:1–14, 1965.

T. Groves and J. O. Ledyard. Optimal allocation of public goods: A solution to the 'free rider' problem. *Econometrica*, 45:783–809, 1977.

H. Gintis, E. A. Smith, and S. Bowles. Costly signaling and cooperation. *Journal of theoretical Biology*, 213:103–119, 2001.

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

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, and J. Shen. Tacit coordinationin contribution-based grouping withtwo endowment levels. *Research in Experimental Economics*, 13:13–75, 2010b.

S. Cabrera, E. Fatas, J. Lacomba, and T. Neugebauer. Splitting leagues: promotion and demotion in contribution-based regrouping experiments. *Experimental Economics*, 16:426–441, 2013.

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

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

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

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

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

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

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.

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

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.

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

I. Kant. *Grundlegung zur Metaphysik der Sitten.* Johann Friedrich Harknoch, 1785.

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.

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

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

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

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

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

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

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? *The American Economic Review*, 87:829–846, 1997.

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.

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. *The American Economic Review*, 100:541–556, 2010.

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

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. 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.

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.

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

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

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

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

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.

J. Bentham. *An Introduction to the Principles of Morals and Legislation.* Clarendon Press, 1907.

J. Rawls. *A Theory of Justice.* Belknap Press, 1971.

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

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

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

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

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

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

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

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.