



**DEPARTMENT OF ECONOMICS**

**DISCUSSION PAPER SERIES**

**WHEN DOES ONE BAD APPLE SPOIL THE BARREL?  
AN EVOLUTIONARY ANALYSIS OF COLLECTIVE ACTION**

**David P. Myatt and Chris Wallace**

Number 269

June 2006

Manor Road Building, Oxford OX1 3UQ

# When Does One Bad Apple Spoil the Barrel?

## An Evolutionary Analysis of Collective Action

David P. Myatt

Department of Economics, University of Oxford

david.myatt@economics.ox.ac.uk

Chris Wallace

Department of Economics and Trinity College, University of Oxford

christopher.wallace@economics.ox.ac.uk

22 June 2006.<sup>1</sup>

**Abstract.** This paper studies  $n$ -player collective-action games in which a public good is produced if and only if  $m$  or more volunteers contribute to it. Quantal-response strategy revisions allow play to move between equilibria in which a team of  $m$  players successfully provide, and an equilibrium in which the collective action fails. A full characterisation of long-run play reveals the determinants of success. These include the correlation between players' costs of provision and their valuations for the good. The addition of an extra "bad apple" player can "spoil the barrel" by destabilising successful teams. A contemporary application is the team-based provision of open-source software. The analysis reveals the features of successful open-source projects, and suggests a rationale for limiting the pool of possible contributors.

### 1. COLLECTIVE ACTION

The problems of collective action have long concerned social scientists. The classic work of Olson (1968) sparked an interest that spanned political science, sociology, and economics. To economists, the Olson view is that a collective action involves the private production of a public good or the private exploitation of a common-pool resource. For this view, a private individual's failure to incorporate externalities results in socially sub-optimal public-good provision or a tragedy of the commons. Political scientists and sociologists have also found many instances of these problems of collective action. These include voter participation (Palfrey and Rosenthal, 1985; Bendor, Diermeier, and Ting, 2003; Diermeier and Van Mieghem, 2005a), interest groups (Oliver, Marwell, and Teixeira, 1985; Oliver and Marwell, 1988; Marwell, Oliver, and Prael, 1988), industry lobbies (Hansen, 1985), consumer boycotts (Diermeier and Van Mieghem, 2005b; Innes, 2006), and many others.

---

<sup>1</sup>**Date Printed.** June 22, 2006 (Collv5.tex). **JEL Classification.** C72, C73, and H41.

**Keywords.** Collective action, evolution, teams, equilibrium selection, concordance, open-source software.

**Acknowledgements.** This paper includes elements of earlier work released under the title "The Evolution of Collective Action." For their helpful comments, thanks are owed to seminar participants at Keele University, Oxford University, and the "Teamwork" conference held at St. Catherine's College, Oxford in 2003. For detailed suggestions, thanks are offered to David Levine, Juuso Välimäki, and the anonymous referees.

Collective-action problems are important for contemporary economics because there are many sectors involving the widespread private provision of public goods. A leading example is the voluntary provision of open-source software. Under the terms of “copyleft” licensing regimes such as the GNU Public Licence (GPL) of the Free Software Foundation, open-source software may be freely redistributed, modified, and used by others.<sup>2</sup> Such software is central to the internet: recent surveys reveal that two-thirds of world-wide-web hostnames are hosted using the open-source Apache web server; the open-source operating system Linux enjoys huge popularity as a server platform; and the open-source web browser Firefox has begun to erode the dominance of Internet Explorer.<sup>3</sup> These are significant components of current information technologies.<sup>4</sup> It is interesting to note that many successful open-source projects (including Apache, Linux, Perl, and Sendmail) are used and produced by computing specialists, whereas leading office tools such as word processors and spreadsheets largely remain within the conventional closed-source world.

Naturally, the open-source movement has attracted the attention of economists. Lerner and Tirole (2002), for instance, viewed the career concerns of programmers as a central issue, proposing that such individuals enhance their future career prospects by writing high-quality open-source applications. Whether their viewpoint is accepted or not, open-source software may still be viewed as a public-good provision problem, albeit one in which (owing to career concerns) the act of provision may sometimes be beneficial rather than costly. Analyses of open-source software as a straightforward public-good provision problem are rare. A notable exception is Johnson (2002), who studied the symmetric equilibrium of a “volunteer’s dilemma” game in which the public good may be produced by a single contributor. However, many open-source software projects rely on the joint participation of a team of contributors.<sup>5</sup>

Since larger projects rely on such teamwork, a collective action may involve problems of coordination. The focus of this paper is a Palfrey-Rosenthal (1984) game in which a public good is provided if and only if a team of  $m$  or more volunteers drawn from a pool of  $n$  players contribute to it. When  $m > 1$ , teamwork is a pre-requisite for success. So long as an individual’s benefit from public-good provision exceeds the private cost of individual participation (this might well be true when  $m$  is large) there are  $\binom{n}{m}$  equilibria in which the public good is provided and a single equilibrium in which no player participates. The equilibrium-selection problem is a “teamwork dilemma” that leads to three central questions. When will the collective action succeed? If it succeeds, who will be the members of the providing team? What role does the size and composition of the contributor pool play?

<sup>2</sup>The GPL is available from <http://www.gnu.org/licenses/>. Developers are free to use and modify software released under the GPL so long as any extensions to it are released under the same licensing terms. The legal foundations of open-source licensing were described by Lerner and Tirole (2005, pp. 22–4).

<sup>3</sup>See the data reported by <http://news.netcraft.com/> and <http://marketshare.hitslink.com/>.

<sup>4</sup>The users of open-source projects include social scientists. For instance, the open-source “R” project (<http://www.r-project.org/>) is a leading platform for statistical computing.

<sup>5</sup><http://sourceforge.net/> is the largest open-source development site. Over 100,000 projects are hosted, and involve over 1,000,000 registered users. The vast majority of these projects involve multiple developers.

The identification of Nash equilibria does not answer these questions.<sup>6</sup> Instead, this paper admits the possibility that play may vary over time. A strategy-revision process is considered in which each player periodically chooses a new strategy (contribute or not) in response to the current state of play. Of course, if strategy revisions are myopic best replies, then the process will lock in to pure-strategy Nash equilibrium states, and the selection problem remains. One possibility, following Kandori, Mailath, and Rob (1993) and Young (1993), would be to allow players to “mutate” against the flow of play by choosing a non-best-reply with some small probability. When a revising player chooses to join a collective action even when not pivotal to its success, the process experiences a low probability “birth.” Similarly, when a revising player chooses to leave a successful  $m$ -strong team of contributors, then the process experiences a “death.” Under a mutation specification, births and deaths share equal probabilities, and could be interpreted as “mistakes” in the revision process.

Here, however, revisions are made via quantal response (McKelvey and Palfrey, 1995).<sup>7</sup> Thus the probability of a birth or a death may depend upon the underlying payoffs.<sup>8</sup> For instance, if a player’s private cost of contribution is low then any benefit from the act of contribution (such as the career concerns that are central to the Lerner-Tirole view of open-source software) may overwhelm it; a birth is more likely. Similarly, a contributing team member is less likely to die when the public good is highly prized. An individual player is characterised by a “birth cost” and a “death cost.” These two variables index the resistance to a birth and death respectively. So, a player with a low birth cost chooses to contribute against the flow of play with relatively high probability. Birth and death costs are determined not only by the payoffs of the game, but also by the relative “noise” in a player’s quantal response. Under the usual random-utility interpretation of a quantal-response specification, a player with particularly variable payoffs will tend to have low birth and death costs.

For a quantal-response strategy-revision specification, the paper characterises the associated long-run distribution over strategy profiles. In particular, when quantal responses approximate myopic best replies the process spends almost all time local to the pure-strategy Nash equilibria of the underlying collective-action game. While particularly sharp results are available for this limiting case, the analysis may also be applied away from the limit.

When will a collective action succeed? This depends upon the ease with which teams form and disband. To build a team from scratch,  $m - 1$  players must take a costly action which yields no immediate benefit; the  $m$ th team member is pivotal and so finds it myopically optimal to participate. Team disruption, on the other hand, takes but one death: a single

<sup>6</sup>Palfrey and Rosenthal (1984) characterised mixed-strategy Nash equilibria of a teamwork dilemma in which a subset of the players randomise. Even when payoffs are symmetric, there are values of  $m$  and  $n$  for which there is no symmetric mixed-strategy equilibrium; a simple “appeal to symmetry” cannot address the equilibrium-selection problem, unless that appeal is content that no player will ever contribute.

<sup>7</sup>This modelling strategy has been exploited in a series of papers by Blume (1995, 1997, 2003) and Blume and Durlauf (2001), who studied logit-driven quantal responses. Using an approach related to the one taken here Diermeier and Van Mieghem (2005b) presented a model of consumer boycotts.

<sup>8</sup>Hence this specification generates “state-dependent mutations” in the sense of Bergin and Lipman (1996).

participant must choose (against the flow of play) to leave a successfully operating team, hence prompting its collapse. With symmetric players, a comparison of these elements reveals that a collective action will succeed when  $m - 1$  births are less costly than a single death.

Who are the members of a successful team? They are (usually) the “most enthusiastic” players who face the lowest birth costs. Such enthusiasts might have low contribution costs. These efficient providers will be least likely to leave a successful team. Alternatively, enthusiasm may stem from payoff variability. A player with relatively noisy payoffs will face not only a low birth cost, but also a low death cost. Such a team member can be a “bad apple” within the team, and cause the collective action to fail. These bad apples (players with low birth and death costs) can also appear when the players with low contribution costs in the contributor pool are also the ones that have low valuations for the public good.

This is not the only instance of the bad-apple effect. Consider a player outside a successfully operating team, with a reasonably low birth cost. If this player also has a low death cost, then such a player might destabilise a successful collective action from without: the player voluntarily joins the team as the  $(m + 1)$ th contributor, an incumbent team member finds it myopically optimal to leave, the “bad apple” subsequently drops out with high probability, and the team collapses. This story illustrates how the wider pool of potential contributors matters. Such players, whilst not playing a major role in long-run provision, nevertheless may supplant more reliable team members, disrupting the team along the way.

Bad-apple effects shed light on the role played by the size and composition of the contributor pool. The success of a collective action is favoured when low birth costs are associated with high death costs (where the index of association is the concordance of a copula) so that enthusiasts are hard to kill. In an open-source software context, this negative association arises when the application is relatively useful to the most talented programmers. A networking utility might fit with this description, whereas a spreadsheet might not.

Bad-apple effects are present even when quantal responses do not approximate best replies. The worst kind of bad apple is a player with an intermediate contribution cost: this rotten apple joins in now and again, but tends not to stick around for the long haul. Welfare increases when such rotten apples are excluded from the pool of potential contributors. Returning to the open-source context once more, welfare is optimised when the contributor pool is restricted to a subset of players with sufficiently low contribution costs.

Section 2 describes the evolving play of a collective-action game. Section 3 presents results for vanishing noise, which are then applied to voluntary open-source software provision in Section 4. Section 5 contains results for non-vanishing noise, and shows why it is socially optimal to restrict the contributor pool. The proofs of the propositions in the main text are relegated to Appendix C. These build upon two central theorems: Theorem 1 (Appendix A) provides a complete characterisation of the ergodic distribution for vanishing noise; Theorem 2 (Appendix B) relates the notions of concordance and team success.

## 2. THE EVOLUTION OF COLLECTIVE ACTION

In this section the collective-action game of interest is described, the mechanism by which play evolves is introduced, and various quantal-response interpretations are considered.

In a simultaneous-move  $n$ -player game, player  $i$  selects  $z_i \in \{0, 1\}$ , where  $z_i = 1$  represents participation in a collective action. A pure-strategy profile  $z \in Z \equiv \{0, 1\}^n$  generates a payoff  $u_i(z)$  for player  $i$ . In the context of the process described below,  $z$  is a “state of play” from the state space  $Z$ . It proves useful to write  $|z| \equiv \sum_i z_i$  for the number of contributors, and  $Z_k \equiv \{z : |z| = k\}$  for the  $k$ th “layer” of the state space.

Further notation proves helpful. Of interest will be the comparison of states that differ by the action of player  $i$ . Starting from  $z$ , write  $z^{i+}$  for the state obtained by setting  $z_i = 1$  and  $z^{i-}$  for the state obtained by setting  $z_i = 0$ ; hence  $z \in \{z^{i-}, z^{i+}\}$ . The “contribution incentive” for player  $i$  is  $\Delta u_i(z) \equiv u_i(z^{i+}) - u_i(z^{i-})$ . With very little loss of generality, and to simplify the exposition, the contribution incentive is assumed to be non-zero for any  $z$ . The set of pure-strategy Nash equilibria is simply  $Z^* = \{z : z_i = 1 \Leftrightarrow \Delta u_i(z) > 0\}$ .

The collective-action games of interest are “ $n$  choose  $m$ ” games for some  $m > 1$ : a player has an incentive to contribute if and only if that player’s participation is pivotal to the formation of a “team” of  $m$  players. Using the contribution-incentive terminology, this means that

$$\Delta u_i(z) > 0 \quad \Leftrightarrow \quad |z^{i+}| = m.$$

For such games, there are  $\binom{n}{m} + 1$  pure-strategy Nash equilibria: the  $\binom{n}{m}$  states of play in the  $m$ th layer  $Z_m$ , plus the single state in  $Z_0$ . The equilibria in  $Z_m$  involve team success: each team member (a player satisfying  $z_i = 1$ ) is pivotal. In contrast, failure involves  $z_i = 0$  for all  $i$ : when  $m > 1$  no individual is able to provide unilaterally. Here “team success” corresponds to the joint provision of a public good. Setting  $v_i > c_i > 0$ , a payoff specification might be

$$u_i(z) = v_i \times \mathcal{I}[|z| \geq m] - z_i c_i,$$

where  $\mathcal{I}[\cdot]$  is the indicator function. Thus player  $i$ ’s private valuation for the public good is  $v_i$ , and the private contribution cost is  $c_i$ . This is the threshold public-good provision game studied by Palfrey and Rosenthal (1984). However, many other collective-action games yield the same best-response structure and the same set of pure-strategy Nash equilibria. Since  $m > 1$ , the important feature is that players face a “teamwork dilemma.”<sup>9</sup>

Rather than study Nash equilibria, attention turns to evolving play. At each time  $t$  the state of play  $z_t \in Z$  is updated via a one-step-at-a-time strategy-revision process: a player  $i$  is randomly selected and chooses an action based solely on the current state, which is then updated to  $z_{t+1} \in \{z_t^{i-}, z_t^{i+}\}$ . This generates a Markov chain on  $Z$ , with transition probabilities  $\Pr[z \rightarrow z'] \equiv \Pr[z_{t+1} = z' \mid z_t = z]$ . The transitions involve steps up and down

<sup>9</sup>In contrast,  $m = 1$  yields a “volunteer’s dilemma” in which there are  $n$  “individual success” equilibria; the issue is no longer whether the collective action will succeed, but rather the identity of the public-good provider. A complementary analysis of this “ $n$  choose 1” case is available from the authors upon request.

between the layers of the state space. A step up is the “birth” of a new contributor, and involves a (myopic) best reply by the revising player if and only if  $z_{t+1} \in Z_m$ ; that is, if and only if the new contributor is the pivotal  $m$ th member of a successful team. Otherwise, a birth follows a revision that is against the flow of play. Similarly, a step down is the “death” of an existing contributor. This is against the flow of play if and only if  $z_t \in Z_m$ .

Allowing a revising player to choose the strict best reply to the current state yields path-dependence; the process will lock in to any pure-strategy Nash equilibrium. Here, players are assumed to choose against the flow of play with some probability. Formally,

$$\Pr[z^{i-} \rightarrow z^{i+}] = \frac{1}{n} \times \begin{cases} 1 - d_i & z^{i+} \in Z_m, \\ b_i & \text{otherwise,} \end{cases}$$

and  $\Pr[z^{i+} \rightarrow z^{i-}] = \frac{1}{n} \times \begin{cases} d_i & z^{i+} \in Z_m, \\ 1 - b_i & \text{otherwise.} \end{cases}$

Player  $i$  will normally participate if and only if a contribution is pivotal. When not, player  $i$  contributes only with some (perhaps small) birth probability  $b_i > 0$ . Similarly, player  $i$  exits a successful team (or equivalently fails to join in when  $m - 1$  others contribute) with some (again, perhaps small) death probability  $d_i > 0$ . These “mutations” allow the strategy-revision process to escape from pure-strategy equilibria and move around the state space. If  $b_i > 0$  and  $d_i > 0$  for each  $i$ , the strategy-revision process is an ergodic Markov chain on  $Z$ , and there exists a unique ergodic distribution over  $Z$  satisfying, for any initial conditions,

$$p_z = \lim_{t \rightarrow \infty} \Pr[z_t = z],$$

so that  $p_z$  reveals how often  $z$  is played in the long run. If play spends most of the time in  $Z_m$ , then the collective action succeeds; if play languishes in  $Z_0$ , then the action fails.

One possible specification would be  $b_i = d_i = \varepsilon > 0$  for an “error” probability  $\varepsilon > 0$ . The mistakes are “state independent” mutations, since the probability of a non-best-reply does not depend upon  $z$ . A standard approach (Foster and Young, 1990; Kandori, Mailath, and Rob, 1993; Young, 1993) would be to examine  $p_z$  as  $\varepsilon \rightarrow 0$ . In the limit, the distribution will place all weight on a “stochastically stable” subset of states; when the stochastically stable set is a single pure-strategy Nash equilibrium then that equilibrium is “selected.”

Here, however a more general “state dependent” model is considered. Birth and death probabilities differ from each other and across players. As “noise vanishes” (that is, as  $\varepsilon \rightarrow 0$ ) these probabilities decline at different rates. Formally,

$$\varepsilon \times \log \left[ \frac{1 - b_i}{b_i} \right] = \beta_i \quad \text{and} \quad \varepsilon \times \log \left[ \frac{1 - d_i}{d_i} \right] = \delta_i. \quad (1)$$

An inspection of (1) reveals that  $\beta_i$  is the (exponential) rate at which  $b_i$  vanishes as  $\varepsilon \rightarrow 0$ ; it is the “birth cost” of a transition made against the flow of play by a new contributor.

Similarly,  $\delta_i$  is the “death cost” of a step down from  $Z_m$ .<sup>10</sup> The specification arises naturally from a model of myopic quantal response. Suppose that a revising player  $i$  chooses a quantal response in the sense popularised by McKelvey and Palfrey (1995), so that

$$\frac{\Pr[z_{i,t+1} = 1]}{\Pr[z_{i,t+1} = 0]} = \exp\left(\frac{\Delta u_i(z_t)}{\varepsilon}\right). \quad (2)$$

Hence the log odds of participating versus not are linear in the contribution incentive. This carries a random-utility interpretation: if a revising player’s contribution incentive is perturbed by a logistic error, then the logit is obtained. Birth and death probabilities respond to the payoffs at stake. For instance, if player  $i$ ’s private cost of participation is small, then the birth cost  $\beta_i$  might also be small. Similarly, if the benefit from a successful collective action is large, then the death cost of an active team member might be correspondingly large. To see the logit quantal-response model in action, take the Palfrey and Rosenthal (1984) threshold public-good provision game, for which  $u_i(z) = v_i \mathcal{I}[|z| \geq m] - z_i c_i$ . Then

$$\Delta u_i(z) = \begin{cases} v_i - c_i & |z^{i+}| = m, \\ -c_i & \text{otherwise.} \end{cases} \quad (3)$$

This specification yields  $\beta_i = c_i$  and  $\delta_i = v_i - c_i$ , which satisfy (1). A logit specification is the focus of Section 5, where long-run play is characterised for any  $\varepsilon > 0$ .

The results of Sections 3 and 4 (which consider the properties of long-run play as  $\varepsilon \rightarrow 0$ ) also apply to a wider class of specifications. For example, strategy revisions are made by probit quantal-response if player  $i$  chooses to contribute if and only if  $\Delta \tilde{u}_i(z) > 0$  where  $\Delta \tilde{u}_i(z) \sim N(\Delta u_i(z), \varepsilon \times \sigma_i^2(z))$ . For the threshold game, set  $\Delta u_i(z)$  as above and

$$\sigma_i^2(z) = \begin{cases} \gamma_i^2 & |z^{i+}| = m, \\ \xi_i^2 & \text{otherwise.} \end{cases} \quad (4)$$

Then the probability that player  $i$  chooses to contribute against the flow of play is given by  $b_i = 1 - \Phi[c_i/(\xi_i \times \sqrt{\varepsilon})]$ , where  $\Phi(\cdot)$  is the cumulative distribution of the standard normal. This birth probability vanishes as the variance in the probit specification, indexed by  $\varepsilon$ , falls to zero. For small  $\varepsilon$ , the birth probability can be approximated by a density. Going into the tails, the log of the density of the normal falls with the square of its argument. So, as  $\varepsilon \rightarrow 0$ ,

$$\varepsilon \log \left[ \frac{1 - b_i}{b_i} \right] \rightarrow \beta_i \quad \text{and} \quad \varepsilon \log \left[ \frac{1 - d_i}{d_i} \right] \rightarrow \delta_i \quad \text{where} \quad \beta_i = \frac{c_i^2}{2\xi_i^2} \quad \text{and} \quad \delta_i = \frac{(v_i - c_i)^2}{2\gamma_i^2}. \quad (5)$$

Other specifications (see Appendix A) also fit into the “birth and death cost” framework.<sup>11</sup> The results of Sections 3 and 4 require only that (1) holds in the limit as  $\varepsilon \rightarrow 0$ , as it does for the probit specification, whereas Section 5 assumes that (1) holds for all  $\varepsilon > 0$ .

<sup>10</sup>If birth and death costs differ, then birth and death probabilities vanish to zero at different rates as  $\varepsilon \rightarrow 0$ . This generates state-dependent mutations in the sense of Bergin and Lipman (1996).

<sup>11</sup>For instance, when  $\Delta \tilde{u}_i(z)$  follows a generalised error distribution (Harvey, 1981) then the associated birth and death costs satisfy  $\beta_i \propto c_i^\nu$  and  $\delta_i \propto (v_i - c_i)^\nu$ , where  $\nu \geq 1$  is a tail-thickness parameter.



### 3. COLLECTIVE ACTION IN THE LONG RUN: TEAM SUCCESS AND FAILURE

Long-run behaviour is characterised by the ergodic distribution  $p_z \equiv \lim_{t \rightarrow \infty} \Pr[z_t = z]$ . This section considers the properties of this distribution as  $\varepsilon \downarrow 0$ ; the case of “vanishing noise.”

When  $\varepsilon$  is small, the flow of play almost always follows the direction of best reply, and hence play tends to “lock in” to the pure-strategy Nash equilibrium states contained in  $Z_0$  (when the collective action fails) or states in  $Z_m$  (when the collective action succeeds).

**Definition.** *Allowing noise to vanish ( $\varepsilon \downarrow 0$ ) teamwork succeeds if and only if  $\sum_{z \in Z_m} p_z \rightarrow 1$ , and otherwise it fails. A specific  $m$ -strong team  $z \in Z_m$  succeeds if and only if  $p_z \rightarrow 1$ .*

Characterisation of the ergodic distribution is straightforward when the players share the same birth and death costs  $\beta$  and  $\delta$ . A strategy revision depends only on the number of contributors in the current state. Hence, to analyse the behaviour of the Markov chain, it is sufficient to keep track of which layer of the state space is currently occupied.

Formally, the strategy-revision process reduces to a simpler Markov chain with a state of  $|z_t| \in \{0, 1, \dots, n\}$ . Writing  $\Pr[Z_k \rightarrow Z_{k+1}] \equiv \Pr[z_{t+1} \in Z_{k+1} | z_t \in Z_k]$ , such a “step up” comprises two elements. A non-contributor is chosen to revise with probability  $(n - k)/n$ . When  $k \neq m - 1$ , participation is not a best reply, and a birth against the flow of play occurs with probability  $b$ . When  $k = m - 1$ , however, the revising player is pivotal to the success of the collective action and hence joins in with probability  $1 - d$ . Summarising,

$$\Pr[Z_k \rightarrow Z_{k+1}] = \frac{n - k}{n} \times \begin{cases} 1 - d & \text{if } k = m - 1, \text{ and} \\ b & \text{otherwise.} \end{cases}$$

Similar considerations lead to the “step down” transitions, which occur with probability

$$\Pr[Z_k \rightarrow Z_{k-1}] = \frac{k}{n} \times \begin{cases} d & \text{if } k = m, \text{ and} \\ 1 - b & \text{otherwise.} \end{cases}$$

The transitions characterise a simple birth-death process on  $\{0, 1, \dots, n\}$ , and the ergodic distribution is tied down by detailed-balance conditions. The long-run relative probability for two neighbouring layers is determined by their relative transition probabilities:  $\Pr[Z_k \rightarrow Z_{k+1}] \times \sum_{z \in Z_k} p_z = \Pr[Z_{k+1} \rightarrow Z_k] \times \sum_{z \in Z_{k+1}} p_z$ . Substituting the step-up and step-down transition probabilities from above, the relative likelihood of team success versus failure is

$$\frac{\sum_{z \in Z_m} p_z}{\sum_{z \in Z_0} p_z} = \frac{n!}{m!(n - m)!} \times \left[ \frac{b}{1 - b} \right]^{m-1} \times \frac{1 - d}{d} = \binom{n}{m} \times \exp\left(\frac{\delta - (m - 1)\beta}{\varepsilon}\right). \quad (6)$$

As  $\varepsilon \rightarrow 0$ , the sign of  $\delta - (m - 1)\beta$  determines whether this ratio explodes or vanishes.<sup>12</sup> Similar calculations confirm that  $\sum_{z \in Z_k} p_z \rightarrow 0$  for other layers  $k \notin \{0, m\}$ . Therefore, a comparison of  $(m - 1)\beta$  and  $\delta$  establishes whether the team is successful.

<sup>12</sup>For  $\lim_{\varepsilon \rightarrow 0} [\sum_{z \in Z_m} p_z / \sum_{z \in Z_0} p_z] = \lim_{\varepsilon \rightarrow 0} [\exp((\delta - (m - 1)\beta)/\varepsilon)]$  to hold requires only that (1) holds in the limit. Hence, while (6) is true only when (1) holds, Proposition 1 holds for limit cases such as (5).

**Proposition 1.** *With symmetric players, teamwork succeeds if and only if  $(m - 1)\beta < \delta$ .*

Intuitively, to build a successful team from scratch requires  $m - 1$  volunteers to choose against the flow of play and contribute to the collective action; for the  $m$ th volunteer, participation is a (myopic) best reply. In contrast, a successful team is disrupted by the exit of a single deviant. Proposition 1 reveals that a state-dependent specification is often crucial to team success. When the “noise” stems from state-independent mutations (constant-probability errors in strategy revisions) then  $\beta = \delta$ , and hence (for  $m > 2$ ) teamwork always fails. A specification  $\beta < \delta$  seems appropriate when the private cost of a contribution is small ( $c$  is small for the Palfrey-Rosenthal game) and the value of success is large ( $v$  is large).

If teamwork succeeds, who will be the members of the successful team? Proposition 1 is mute: given symmetry, all teams in  $Z_m$  are equally likely. Asking who participates is interesting only when players differ in some way. Hence, to take up the second question, attention turns to an environment in which players’ birth and death costs (generically) differ.<sup>13</sup>

Without loss of generality, players will be labelled in birth-cost order:  $\beta_1 < \beta_2 < \dots < \beta_n$ . Thus a player with a lower label  $i$  is, heuristically, more enthusiastic. Similarly, for states in  $Z_m$ , the team  $z^\dagger \in Z_m$  in which  $z_i = 1 \Leftrightarrow i \leq m$  might be called the most enthusiastic team.

When players differ, it is insufficient to track the layer in which the state resides: the identity of contributors is important. Detailed-balance conditions do not hold in general, and global-balance conditions are not amenable to direct analysis.<sup>14</sup> The literature offers two techniques for the characterisation of the ergodic distribution as noise vanishes. Both begin by identifying the limit sets of a noiseless strategy-revision process; for this paper, this is when strategy revisions are myopic best replies. Such a limit set is a subset of communicating states from which the (noiseless) Markov chain cannot escape. For collective-action games, the limit sets are the  $\binom{n}{m} + 1$  singletons corresponding to pure-strategy Nash equilibria.<sup>15</sup>

The “tree surgery” technique introduced by Foster and Young (1990) was popularised by Kandori, Mailath, and Rob (1993) and Young (1993). The analyst builds directed graphs on the space of limit sets that form “trees” leading to a single “root” limit set. A branch of a tree corresponds to the least-resistant path between two limit sets, and its “resistance” (for the purposes of the present paper) is the sum of any birth and death costs encountered along the way. The resistance of a tree is the sum of the resistances of its component branches. If a tree rooted at a limit set has strictly lower resistance than any other tree rooted at any other limit set, then it attracts all probability in the ergodic distribution as noise vanishes.<sup>16</sup>

<sup>13</sup> $\beta_i \neq \beta_j$  and  $\delta_i \neq \delta_j$  for all  $i \neq j$ . These requirements are satisfied generically; they fail only on a set of Lebesgue measure zero when the parameter space is, for instance, an appropriate subset of  $\mathcal{R}^{2n}$ . When teamwork does not succeed and either  $m \geq 3$  or  $m = 2$  and  $\delta_1 > \beta_1$ , generically  $\sum_{z \in Z_0} p_z \rightarrow 1$  as  $\varepsilon \rightarrow 0$ .

<sup>14</sup>There is, however, a class of configurations for which detailed-balance conditions hold; when  $\beta_i + \delta_i$  is constant across the set players. This case admits analysis for non-vanishing noise. See Section 5.

<sup>15</sup>The Nash equilibria in  $Z_0$  and  $Z_m$  are absorbing states, and hence automatically limit sets. Since the collective-action games studied here are acyclic, there are no multiple-member limit sets.

<sup>16</sup>The limit set in question is “stochastically stable” and its corresponding Nash equilibrium is “selected.”

The alternative is the radius-coradius technique of Ellison (2000). The “radius” of a limit set is the difficulty of escaping from it, whereas the “modified coradius” is the difficulty of returning to it from elsewhere in the state space. If the radius of a limit set exceeds its modified coradius, it attracts all probability in the ergodic distribution as noise vanishes.

Whereas the radius-coradius approach only provides sufficient conditions for the stochastic stability of a limit set (and hence “selection” of a Nash equilibrium) tree surgery also provides necessary conditions. Theorem 1 (Appendix A) uses tree surgery to characterise completely the stochastically stable sets of evolving collective action, and hence determines which team succeeds and when. A shortcoming is that the mechanics of tree surgery often fail to reveal why particular limit sets are stochastically stable. The radius-coradius technique is more helpful in this regard, since its components are respectively measures of the persistence of a Nash equilibrium and its attractiveness. For a collective action, these factors measure the difficulty of building a team from scratch and the barrier to the disruption of a successful team and its subsequent collapse back to failure. Overall, this means that the radius-coradius approach can help provide insight, although the formal results emerging from it are sometimes limited.<sup>17</sup> Based on these observations, this paper takes a somewhat eclectic approach. Whereas the propositions stated below are derived as corollaries of the aforementioned Theorem 1, the intuition developed throughout the text (formal proofs are relegated to Appendix C) builds upon the radius-coradius methodology.

It proves useful to divide collective-action games into three different classes depending upon the relationship between birth and death costs across the set of players.

The first case is when more enthusiastic players (with lower birth costs) are more reluctant to disrupt a successfully operating team (they have high death costs). Given that players are labelled in birth-cost order (so that  $\beta_1 < \dots < \beta_n$ ) this obtains when death costs are decreasing:  $\delta_1 > \dots > \delta_n$ . The members of the most enthusiastic team  $z^\dagger$  are the hardest to kill, and so the candidates for stochastic stability are  $z^\dagger$  and the team-failure state in  $Z_0$ .

To calculate the persistence of team failure, consider the “basin of attraction” of  $Z_0$ . This is the set of states from which a noiseless process (myopic best reply) will always descend down into  $Z_0$ ; it consists of layers  $Z_{m-2}$  and below. To escape from this basin, beginning from team

<sup>17</sup>A desire for economic insight perhaps sparked Ellison’s remark (2000, p. 18) that the tree-surgery algorithm

“... is sufficiently complex as to have made the whole literature seem a bit mysterious—a commonly expressed frustration is the feeling that the typical paper writes down a model, says something about trees and after several pages of calculations gives an answer without letting one see how the answer is connected to the assumptions of the model.”

This paper sympathises with his position. Ellison (2000, p. 19) viewed the radius-coradius alternative as

“... sufficiently powerful so as to allow virtually all of the identifications of long-run stochastically stable sets found in the previous literature to be rederived as corollaries.”

His statement certainly applies to early applications of stochastic evolution, such as the selection of risk-dominant equilibria in  $2 \times 2$  coordination games. For collective-action games, however, there are various specific (and plausible) circumstances in which the tree-surgery approach can provide predictions whereas the radius-coradius analysis cannot; a specific example is provided in Footnote 18 just below.

failure a successful team must be built from scratch, requiring  $m - 1$  players to voluntarily join the collective action. The easiest way to do this is for the  $m - 1$  most enthusiastic players to volunteer, at a total birth cost of  $\sum_{i=1}^{m-1} \beta_i$ . Using Ellison's (2000) terminology, this is the *radius* of  $Z_0$ . This term also indexes the attractiveness of  $z^\ddagger$ . Intuitively, the team-failure state is "furthest away" from  $z^\ddagger$ . The easiest way to return back to  $z^\ddagger$  is for the  $m - 1$  most enthusiastic players to volunteer, so that  $\sum_{i=1}^{m-1} \beta_i$  provides an upper bound to how far away the process can go; using Ellison (2000) terminology once more, this is the *coradius* of  $z^\ddagger$ . Taken together, the  $m - 1$  lowest birth costs measure the attractiveness of team success and the persistence of team failure. However, a determination of the stochastically stable set also requires consideration of the persistence of success and the attractiveness of failure.

The attractiveness of failure depends upon the ease with which a successful team collapses. Beginning from the most enthusiastic team  $z^\ddagger$ , one way to collapse is for one of the team members to die. The team member with the lowest death cost is player  $m$ , and hence a death cost of  $\delta_m$  allows the process to move down to layer  $Z_{m-1}$ . From there, a myopic best-reply by a contributor will lead to a further step down into the basin of attraction of  $Z_0$ . Hence  $\delta_m$  is an upper bound to the distance of  $z^\ddagger$  from  $Z_0$ , and is a candidate for the coradius of  $Z_0$ . Certainly, if  $\sum_{i=1}^{m-1} \beta_i > \delta_m$  then the collective action will fail.

There can be easier ways for the team to collapse. Beginning from  $z^\ddagger$ , suppose that player  $j > m$  volunteers to join the collective action (overcoming a birth cost  $\beta_j$ ) and becomes the  $(m + 1)$ th contributor. If player  $i \leq m$  subsequently encounters a strategy-revision opportunity, a myopic best reply will bring the process back down into some state  $z \in Z_m$ ; starting from the most enthusiastic team, player  $i \leq m$  has been replaced by player  $j > m$ . Now suppose that player  $j$  is chosen to revise once more, and then dies (overcoming a death cost  $\delta_j$ ) so that the process steps down to  $Z_{m-1}$ . If  $\beta_j + \delta_j < \delta_m$ , then this indirect route is an easier path of escape from  $z^\ddagger$ . The quickest indirect route arises when  $j > m$  is chosen to minimize  $\beta_j + \delta_j$ . Combining these observations, the coradius of  $Z_0$  is  $\min[\delta_m, \min_{j>m}[\beta_j + \delta_j]]$ .

An even tighter bound on the attractiveness of team failure can be obtained. The "replace  $i$  with  $j$ " route passes through an intermediate limit set (that is, the state  $z \in Z_m$  discussed above) of the noiseless myopic best-reply process. This means that the birth and death of player  $j$  do not need to happen consecutively; the process can remain at  $z$  for some time and wait for  $j$  to be offered the revision opportunity. Ellison's *modified coradius* incorporates this "step by step" evolution; the appropriate adjustment to the resistance of the path is to deduct the radius of the intermediate limit set. Now, the radius of  $z$  is the easiest way out. This is either the easiest death, which must be player  $j$  and has a death cost of  $\delta_j$ , or the birth of the most enthusiastic player from outside  $z$ . This outside enthusiast must, of course, be player  $i$ : the team member from the most enthusiastic team whom player  $j$  supplanted. Putting these observations together,  $z$  has a radius of  $\min[\delta_j, \beta_i]$  and the modified resistance of the step-by-step route is  $\beta_j + \delta_j - \min[\delta_j, \beta_i]$ . This expression is minimised when the radius of the intermediate step is large. This is so when  $i = m$ , so that  $j$  supplants the

least enthusiastic member of the most enthusiastic team. Choosing player  $j$  to minimise this expression generates the easiest indirect route out of  $z^\dagger$ , with a total resistance of

$$\delta^\dagger \equiv \min_{j>m} [\beta_j + \delta_j - \min[\delta_j, \beta_m]].$$

Comparing direct and indirect routes, the modified coradius of  $z^\dagger$  is simply  $\min[\delta_m, \delta^\dagger]$ . Following from this, a sufficient condition for team failure (which can be derived from the radius-coradius theorem) is  $\sum_{i=1}^{m-1} \beta_i > \min[\delta_m, \delta^\dagger]$ . In fact, this criterion is also necessary.

**Proposition 2.** *When death costs are decreasing (enthusiastic players are the hardest to kill) the most enthusiastic team succeeds if  $\sum_{i=1}^{m-1} \beta_i < \min[\delta_m, \delta^\dagger]$ ; otherwise teamwork fails.*

Proposition 2 reveals two “bad apple” effects. One possible disruption to the collective action is the departure of a member of  $z^\dagger$ ; this player is a bad apple who spoils the barrel from within. To soften this effect, the death costs of members should be high. This is why it makes sense for player  $m$  (with a death cost higher than all players  $j > m$ ) to form the final member of the successful team, even though only the  $m - 1$  most enthusiastic players are required for its construction. When  $\delta_m < \delta^\dagger$ , the bad apple (player  $m$ ) within the team is the weakest link. In contrast, when  $\delta_m > \delta^\dagger$  the bad apple enters from outside the team: player  $j$  enters and becomes the superfluous  $(m + 1)$ th contributor, enabling the myopic departure of an incumbent. It is the subsequent death of player  $j$  that causes the final collapse.

When the bad-apple effect stems from outside the team, player  $j > m$  is not an established member of a successfully operating team, and yet is a critical determinant of that team’s success or failure. Thus, an answer to “when will the collective action succeed?” must begin by evaluating the characteristics of those who will eventually free ride on any success.

Proposition 2 deals with enthusiasts who are hard to kill. A second case is when enthusiasts are easy to kill, so that  $\delta_1 < \dots < \delta_n$ . The most enthusiastic team is also the flakiest; the easiest way to disrupt  $z^\dagger$  is for the most enthusiastic player (with the lowest birth cost  $\beta_1$  and the lowest death cost  $\delta_1$ ) to drop out, and so the bad apple always spoils the team from within. This suggests that the collective-action fails if and only if  $\sum_{i=1}^{m-1} \beta_i < \delta_1$ .

Proposition 3 verifies this hypothesis. However, care must be taken to ensure that those that provide (if indeed they do) are enthusiasts. The reason is that enthusiasts create the bad-apple-inside effect that disrupts teamwork. One possibility, then, would be to replace the flakiest team member with a less enthusiastic but more reliable substitute. Consider, for instance, a team  $z \in Z_m$  identical to  $z^\dagger$  except for the fact that player  $j > m$  supplants player 1. The cost of building  $z$  from scratch is  $\sum_{i=2}^m \beta_i$ , a net increase of  $\beta_m - \beta_1$  relative to the cost of building  $z^\dagger$ . However, the death cost of the bad apple within the team rises to  $\delta_2$ . So, when  $\delta_2 - \delta_1 > \beta_m - \beta_1$ , is  $z$  a better candidate than  $z^\dagger$  for long-run success? Unfortunately not. The reason is that whenever this trick might work, it is undone by the fact that the quickest way out of  $z$  and back to team failure is for player 1 to jump back in

the team and then die. What this means is that  $z$  is more robust to disruption only when there is no hope for team success in the first place. Proposition 3 summarises.<sup>18</sup>

**Proposition 3.** *When death costs are increasing (enthusiastic players are the easiest to kill) the most enthusiastic team succeeds if  $\sum_{i=1}^{m-1} \beta_i < \delta_1$ ; otherwise teamwork fails.*

Propositions 2 and 3 deal with two extreme cases. In the former, birth and death costs are perfectly negatively related: a higher birth cost is always associated with a lower death cost. In the latter, the reverse is true. Negative correlation might be seen as helpful for team success since the bad apple is the player with the  $m$ th-highest death cost. In contrast, with positive correlation the bad-apple effect occurs via the lowest death cost. Of course, in general, the death costs may be arbitrarily ordered in relation to the birth costs.

For the general case, note first that the  $m - 1$  lowest birth-cost players must still play a role.<sup>19</sup> Any team that excludes these players certainly costs more to build. As a result negative correlation between birth and death costs, raising the death costs of these  $m - 1$  players, can only be good for team success. Second, consider a bad apple ( $j \geq m$ ) resident outside a team. As a route via which the team is destabilised, this exit is less effective whenever, for a fixed death cost, the birth cost is higher; or vice versa. Again, heuristically at least, negative correlation between birth and death costs has a positive impact upon team success.

Some language assists a more formal discussion: denote an ordering of death costs  $\delta = (\delta_i)_{i=1}^n$ . Fix the (ordered) values of the birth costs. A new ordering (or “shuffle”) of the same death costs  $\hat{\delta}$  “favours team success” if whenever a team succeeded under  $\delta$  a (possibly different) team certainly will succeed under  $\hat{\delta}$ .<sup>20</sup> Different orderings of death costs retain the marginal distributions of birth and death costs, but change the joint distribution of the rank of birth and death costs. The joint distribution of these ranks is the empirical copula:<sup>21</sup>

$$C(x, y) \equiv \sum_{i=1}^n \mathcal{I}[\beta_i \leq \beta_{(x)}] \times \mathcal{I}[\delta_i \leq \delta_{(y)}] = \sum_{i=1}^x \mathcal{I}[\delta_i \leq \delta_{(y)}],$$

where  $\beta_{(i)} (= \beta_i)$  is the  $i$ th lowest birth cost, and similarly for  $\delta_{(i)}$ . Different orderings of death costs correspond to different copulae. It remains to define a measure of association.

**Definition.**  $C$  is more concordant than  $\hat{C}$  if and only if  $C(x, y) \geq \hat{C}(x, y)$  for all  $x$  and  $y$ .

Concordance provides a (partial) ordering over copulae. Equivalently, since birth costs are arranged in size order by assumption, it is a partial ordering over death-cost shuffles. In

<sup>18</sup>The increasing-death-cost case illustrates the limitations of Ellison’s (2000) method. Suppose that  $\delta_1 > \sum_{i=1}^{m-1} \beta_i > \beta_{m+1}$ . The radius of  $z^\ddagger$  is  $\beta_{m+1}$ . The modified coradius is (at least)  $\sum_{i=1}^{m-1} \beta_i$ . This is larger than radius, and a direct application of Ellison’s method is impossible; yet (Proposition 3) the team  $z^\ddagger$  succeeds.

<sup>19</sup>The  $m - 1$  most enthusiastic players will always be members of any successful team, following from the argument preceding Proposition 3. This claim is formally verified as part of Theorem 1 (Appendix A) which provides a full characterisation of the ergodic distribution for vanishing noise for any death-cost ordering.

<sup>20</sup>A shuffle of  $\delta$  is an ordering  $\hat{\delta}$  such that for each  $i$ ,  $\hat{\delta}_i = \delta_j$  for some  $j$ , (and vice versa).

<sup>21</sup>Copulae usually capture the dependence of continuous variables; see Nelsen (2006) and Cherubini, Luciano, and Vecchiato (2004) for an introduction and applications to finance, respectively. In some recent studies (Mayor, Suñer, and Torrens, 2005, for example) the empirical copula was described as a discrete copula.

fact, the set of such orderings forms a lattice with maximal and minimal members,  $C(x, y) = \min[x, y]$  and  $C(x, y) = \max[0, x + y - n]$  respectively.<sup>22</sup> The former corresponds to  $\delta_{(i)} = \delta_i$  for all  $i$  (perfect concordance) and the latter to the case when birth and death costs are perfectly discordant. An increase in concordance implies an increase in the standard empirical measures of association, such as Spearman's  $\rho$  and Kendall's  $\tau$ .<sup>23</sup> Intuitively a more concordant ordering shifts low death costs towards the players with low birth costs.

**Proposition 4.** *A decrease in concordance favours team success. Hence decreasing death costs are most favourable to success, and increasing death costs are least favourable. A sufficient condition for success is  $\sum_{i=1}^{m-1} \beta_i < \delta_{(1)}$ ; a necessary condition is  $\sum_{i=1}^{m-1} \beta_i < \delta_{(n-m+1)}$ .*

Section 4 provides examples of these four propositions at work. Note, however, that the population size  $n$  has had little impact upon the results: for vanishing noise, it is only the *minimum* cost routes in and out of teams that determine selection, and therefore only *particular* individuals play any role. For non-zero noise, however, there is high probability that any player may join or leave a functioning team, and the population plays a more evident part in evolving play. This is the focus of Section 5; bad-apple effects remain critical.

#### 4. ON THE SUCCESS OF A COLLECTIVE ACTION

Propositions 1–4 take as primitive the somewhat abstract pattern of birth and death costs. This section turns to the more focused setting of a threshold public-good provision game, and explores applications with reference to voluntary open-source software development.

If players value the public good in the same way (they have symmetric valuations) then a threshold public-good provision game (Palfrey and Rosenthal, 1984) arises from the specification  $u_i(z) = v \times \mathcal{I}[|z| \geq m] - z_i c_i$ , where players are ordered so that  $0 < c_1 < \dots < c_n < v$ . As noted in Section 2, this specification generates a contribution incentive

$$\Delta u_i(z) = v \times \mathcal{I}[z^{i+} \in Z_m] - c_i.$$

Given that  $c_i < v$ , it is privately optimal for player  $i$  to participate if  $m - 1$  others do so. This means that it is socially optimal for the collective action to succeed, although looking across the  $m$ th layer  $Z_m$ , players' preferences differ over team-success equilibria; each would rather free ride than provide. When welfare is the utilitarian sum of payoffs, it is socially optimal for the  $m$  most efficient players (with the lowest contribution costs) to provide.

Despite the social optimality of success, teamwork may fail since an individual player is unable unilaterally to provide the public good; this is the essence of the teamwork dilemma. If an individual were able unilaterally to provide, would the collective action succeed? To answer, consider what would happen if an individual were able to compensate others for

<sup>22</sup>These are known in the literature as the minimum and Łukasiewicz discrete copulae respectively. They correspond to the Fréchet-Hoeffding lower and upper bounds for continuously distributed variables.

<sup>23</sup>For instance, Kendall's  $\tau$  measures the incidence of concordant pairs (Nelsen, 2006).

their private costs, but unable to capture the positive externalities from provision. The most efficient way to provide would be to hire the  $m$  most efficient players at a total cost of  $\sum_{i=1}^m c_i$ . Doing so could then be in the interests of a private individual so long as  $v > \sum_{i=1}^m c_i$ .

**Definition.** *A symmetric-valuation collective action is privately feasible  $\Leftrightarrow v > \sum_{i=1}^m c_i$ .*

When a collective action is privately feasible the coordination problem is the source of any team failure; after all, an individual, responding to private benefits and costs alone might profitably provide if only able to do so. In contrast, when the collective action is not privately feasible there is an additional barrier to success: private costs exceed private benefits, and hence the lack of internalisation of positive externalities works against public-good provision.

The criterion for private feasibility is easiest to satisfy when the public good is valuable and contribution costs are low. For instance, a shift down in the distribution of costs, in the usual sense of first-order stochastic dominance, favours private feasibility. In fact, since it is the sum of the lowest costs that is important, an increase in the riskiness of the distribution of costs, in the sense of second-order stochastic dominance, will also favour private feasibility.

Under logit quantal-response, private feasibility is related to the stochastic stability of the most enthusiastic team. From Section 2, birth and death costs satisfy  $\beta_i = c_i$  and  $\delta_i = v - c_i$ . The most enthusiastic team is also the most efficient team. Furthermore, a player with a low cost of contribution (and hence a low birth cost) has a high death cost; enthusiasts are efficient and hard to kill. This scenario fits the “decreasing death cost” case of Proposition 2, and so the most enthusiastic (and efficient) team succeeds if and only if  $\sum_{i=1}^{m-1} \beta_i < \min[\delta_m, \delta^\dagger]$ . When  $\delta_m < \delta^\dagger$ , so that the easiest way to disrupt the most enthusiastic team is via the direct death of its flakiest member, the team-success criterion becomes  $\sum_{i=1}^{m-1} c_i < v - c_m$ , or equivalently  $\sum_{i=1}^m c_i < v$ . This, of course, is the criterion for private feasibility.

It turns out that private feasibility is also the right criterion when the bad apple who spoils the barrel enters from outside the most enthusiastic team. Notice that  $\beta_j + \delta_j = v$ , and so

$$\delta^\dagger = \min_{j>m} [v - \min[v - c_j, c_m]] = \max[c_{m+1}, v - c_m] \Rightarrow \min[\delta_m, \delta^\dagger] = v - c_m.$$

Hence the effect of a bad apple from outside the team can never be strong enough to disrupt the team any more than its own least-enthusiastic member. Summarising:

**Proposition 5.** *Consider a threshold public-good provision game with symmetric valuations. When play evolves via logit quantal response, the most enthusiastic team succeeds if and only if the collective action is privately feasible. Team success is favoured by (i) a shift down in the distribution of costs, and (ii) an increase in the riskiness of the distribution of costs.*

For open-source software, Proposition 5 reveals characteristics of a programming community that assist success: not only is it helpful for costs to be low, it is also helpful if the abilities of potential contributors are varied. A project’s success hinges upon the abilities of the  $m$  best programmers, and variation in ability helps to make the best even better.



Proposition 5 restricts to a scenario in which players share the same valuation for the public good. In the open-source context, programmers may have different uses for a software application. Variation in  $v_i$  can generate additional heterogeneity in the death costs. The implications depend upon the relationship between contribution costs and valuations.

Suppose that the public good is highly valued by enthusiasts. For open-source software, this might be true of a networking utility; such a software application is useful for the skilled programmers who are best able to contribute to its development. A negative relationship between valuation and cost ensures that enthusiasts have high death costs. Since they are most likely to form a successful team, this pattern of valuations and costs tends to minimise disruption and so favours success. Alternatively, the public good may be of little use to enthusiasts. This might be so when the software application is a word processor; skilled programmers (the enthusiasts) might prefer a simple text editor. This may lead to concordant birth and death costs, and hence (following Proposition 4) this favours failure.

Proposition 5 reveals that a riskier distribution of contribution costs helps private feasibility and therefore team success. But what of the distribution of valuations? Consider a world in which contribution costs are negatively related to valuations, so that enthusiastic players highly value the public good. The bad apple in a successfully operating team of enthusiasts is player  $m$ , with the  $m$ th highest valuation and a death cost of  $\delta_m = v_m - c_m$ . If  $m < n/2$  then this player's valuation is above average, in the sense that it is above the median valuation. An increase in the riskiness of the valuation distribution, in the sense of a median-preserving spread, will push up  $v_m$  and therefore raise the death cost  $\delta_m$ ; this favours team success. Of course, if  $m > n/2$ , then the opposite logic applies.

**Proposition 6.** *Consider a threshold public-good provision game where contribution costs are negatively related to valuations, and where play evolves via logit quantal response. An increase in the riskiness of the valuation distribution, in the sense of a median-preserving spread, will favour team success if  $m < n/2$ , but will favour team failure if  $m > n/2$ .*

Propositions 5 and 6 are concerned with distributions looking across the set of players; stealing terminology from econometrics, this is the “between” dimension of the panel of players. Whereas the average valuation and cost may vary across the player set, from the random-utility interpretation of a quantal-response the realised valuation and contribution cost of an individual player may vary over time; this is the “within” dimension of the panel.

The effects of “within” variation are neatly illustrated using a probit specification. Let all players share the same average payoffs, so that  $\Delta u_i(z) = c$  or  $\Delta u_i(z) = v - c$  depending on whether player  $i$  is pivotal to the collective action. Next, set  $\gamma_i^2 = \xi_i^2$  for all  $i$ . Then,

$$\beta_i = \frac{c^2}{2\xi_i^2} \quad \text{and} \quad \delta_i = \frac{(v - c)^2}{2\xi_i^2}.$$

Notice that birth and death costs are both inversely related to the variance parameter  $\xi_i^2$ . Hence players with the lowest birth costs are enthusiastic not because they are efficient

but rather because their payoffs are particularly idiosyncratic; a birth-cost ordering leads to  $\xi_1^2 > \dots > \xi_n^2$ . This means that enthusiasts have the lowest death costs, and so are the flakiest players. Following Proposition 4, such a positive relationship between birth and death costs favours the failure of the collective action.<sup>24</sup> Furthermore, the presence of a player with a particularly high payoff variance can result in team failure. In the open-source context, projects may collapse when they attract the interest of particularly noisy individuals.

For the final result of this section, the focus returns to symmetric players, so that (Proposition 1) the collective action succeeds if and only if  $(m-1)\beta < \delta$ . Under a probit specification, the death cost  $\delta$  of a player is determined by the variance of  $\tilde{v}_i - \tilde{c}_i$ . This, in turn, affected by any “within” correlation between valuation and contribution cost. For instance, if

$$\begin{bmatrix} \tilde{c}_i \\ \tilde{v}_i \end{bmatrix} \sim N \left( \begin{bmatrix} c \\ v \end{bmatrix}, \varepsilon \times \begin{bmatrix} \xi^2 & \rho\xi\varpi \\ \rho\xi\varpi & \varpi^2 \end{bmatrix} \right) \Rightarrow \delta = \frac{(v-c)^2}{2 \times [\xi^2 + \varpi^2 - 2\rho\xi\varpi]},$$

then  $\rho$  is the correlation coefficient between the random-utility valuation and contribution-cost realisations of a player. An increase in this correlation increases the death cost  $\delta$ , while leaving the birth cost unchanged, and hence favours the success of the collective action.

**Proposition 7.** *Consider a symmetric threshold public-good provision game where play evolves via probit quantal-response. An increase in the correlation in the cost and valuation realisations of each individual reduces death costs and so favours team success.*

Intuitively, a team member “dies” whenever  $\tilde{v}_i - \tilde{c}_i < 0$ . This can happen when a player’s contribution cost experiences a positive shock. However, if that player’s contribution cost and public-good valuation are positively correlated then this will be offset by a likely positive shock to the valuation. Positive correlation reduces the “within” variability of the difference between valuation and cost for a player, and so enhances the stability of a successful team.

There is a contrast between the messages emerging from Proposition 4 and 7. Looking across the player set (the “between” dimension of a panel) negative correlation helps to create the desirable discordance between birth and death costs; enthusiasts are reliable, and this favours team success. In contrast, looking at the payoff realisations of an individual player across time (the “within” dimension) negative correlation favours team failure. Hence, when thinking about correlation, it is important to distinguish the “within” and “between” dimensions.

This section closes by turning back once more to the application of open-source software development. Johnson (2002) modelled voluntary software development as an incomplete-information “volunteer’s dilemma” binary-action game in which only a single contribution

<sup>24</sup>This argument, based upon the heteroskedasticity of players’ quantal responses, may also be made using a logit quantal-response specification. To do this, modify (2) so that

$$\frac{\Pr[z_{i,t+1} = 1]}{\Pr[z_{i,t+1} = 0]} = \exp \left( \frac{\Delta u_i(z)}{\lambda_i \times \varepsilon} \right), \quad \text{where} \quad \Delta u_i(z) = \begin{cases} v - c & |z^{i+}| = m, \\ -c_i & \text{otherwise.} \end{cases}$$

The parameter  $\lambda_i$  then indexes the idiosyncrasy of player  $i$ ’s noisy payoffs.

is required for provision of the public-good. Rather than address the equilibrium-selection problem, he instead provided a careful characterisation of the symmetric Bayesian-Nash equilibrium of his game. Whilst it may be argued that an emphasis on such a symmetric equilibrium lacks justification, nevertheless the comparative-static results offered by Johnson (2002) certainly resonate with some of the results of this paper. He observed:

“A puzzle in the open source community is why some obviously useful software does not get written ... while open source word processors and spreadsheets do exist, it is fair to say that only recently have they begun to be comparable in quality to, for example, Microsoft Office. On the other hand, hundreds of other free utilities and applications exist.” (Johnson, 2002, p. 656)

His explanation was that a networking utility might exhibit a negative correlation between value and cost, whereas work processor might exhibit a positive correlation. For evolving play, these two examples correspond to discordance and concordance in birth and death costs, respectively. Proposition 4 provides, therefore, a complementary resolution to the open-source puzzle. However, Proposition 7 suggests caution, since when the “within” dimension is considered, team-success is favoured by positive rather than negative correlation.

## 5. WELFARE AND THE CONTRIBUTOR POOL

This section studies evolving play when noise is bounded away from zero, so that decisions arise from genuine quantal responses rather than near best-replies. Since all states are visited with positive probability, the concepts of “team success” and “team failure” lose meaning and the equilibrium-selection results lying behind Propositions 1–7 are less useful.

Investigating the properties of evolving play when noise is non-vanishing is desirable because the stochastic-stability results presented in earlier sections omit important factors. For instance, and as noted in Section 3, the size  $n$  of the contributor pool plays no role, because only the easiest routes in and out of limit sets matter. However, if the quantal-response noise  $\varepsilon$  is fixed and  $n$  is allowed to rise, then an inspection of (6) reveals that the relative likelihood of team success versus failure can grow without bound, even when  $\delta < (m - 1)\beta$ : there is a larger pool of potential contributors who can help to build a successful team.

Fortunately, progress can be made for  $\varepsilon > 0$ . Consider a symmetric-valuation threshold public-good provision game in which play evolves via logit quantal response, so that

$$\frac{\Pr[z_{i,t+1} = 1]}{\Pr[z_{i,t+1} = 0]} = \exp\left(\frac{\Delta u_i(z_t)}{\varepsilon}\right) \quad \text{where} \quad \Delta u_i(z) = \begin{cases} v - c_i & |z^{i+}| = m, \\ -c_i & \text{otherwise.} \end{cases}$$

This symmetric-valuation specification (players share the same valuation  $v_i = v$  for the public good) leads to  $\beta_i = c_i$  and  $\delta_i = v - c_i$ , so that  $\beta_i + \delta_i = v$ . Birth and death costs are perfectly negatively correlated, and so perfectly discordant. Applying Proposition 2, the most

enthusiastic (and efficient) team succeeds if and only the collective action is privately feasible. The notion of private feasibility is also useful when quantal-response noise is retained.

**Definition.** *For a symmetric-valuation threshold public-good provision game evolving via logit quantal response, the private feasibility of a state  $z$  is  $\psi(z) = v \times \mathcal{I}[|z| \geq m] - \sum_{i=1}^n c_i$ .*

This index of private feasibility comprises a private valuation for the public good (if it is provided) minus the sum of any private costs. Observe that  $\psi(z^{i+}) - \psi(z^{i-}) = u_i(z^{i+}) - u_i(z^{i-}) \equiv \Delta u_i(z)$ , so that the contribution incentive is the effect of a player's participation on the private feasibility of provision. This means that  $\psi(z)$  is an exact potential function (Monderer and Shapley, 1996) and the collective action is an (exact) potential game.

When the play of an exact potential game evolves via logit quantal response, the relative likelihood of jumping back and forth between two neighbouring states is determined by the difference in their potentials. For a binary-action game with an exact potential,

$$\varepsilon \times \log \frac{\Pr[z^{i-} \rightarrow z^{i+}]}{\Pr[z^{i+} \rightarrow z^{i-}]} = \Delta u_i(z) = \psi(z^{i+}) - \psi(z^{i-}). \quad (7)$$

The potential of a state pins down its ergodic probability (Blume, 1997). Formally, consider a distribution  $p_z \propto \exp[\psi(z)/\varepsilon]$ . This is in detailed balance for all pairs of states, so that  $p_z \times \Pr[z \rightarrow z'] = p_{z'} \times \Pr[z' \rightarrow z]$  for all pairs of states.<sup>25</sup> To see this, note that for states that do not directly communicate (they differ by more than one action) the transition probabilities are zero, and so the detailed-balance condition is trivially satisfied. Suppose instead that  $z$  and  $z'$  differ only by the action of  $i$ , so that  $z = z^{i-}$  and  $z' = z^{i+}$ . Then,

$$\frac{p_{z'}}{p_z} = \exp \left[ \frac{\psi(z') - \psi(z)}{\varepsilon} \right] = \exp \left[ \frac{u_i(z') - u_i(z)}{\varepsilon} \right] = \frac{\Pr[z^{i-} \rightarrow z^{i+}]}{\Pr[z^{i+} \rightarrow z^{i-}]},$$

where the final equality follows from the application of (7) above, and so the detailed-balance condition is satisfied once more. This ensures that  $p_z$  is the unique ergodic distribution.

**Proposition 8.** *When birth and death costs are inversely related ( $\beta_i + \delta_i = \bar{\delta} \forall i$ ) then*

$$p_z = \frac{\exp[\psi(z)/\varepsilon]}{\sum_{z' \in Z} \exp[\psi(z')/\varepsilon]} \quad \text{where} \quad \psi(z) = \bar{\delta} \times \mathcal{I}[|z| \geq m] - \sum_{i=1}^n \beta_i z_i.$$

*When play of a symmetric-valuation public-good provision game evolves via logit quantal response (so that  $\beta_i = c_i$ ,  $\delta_i = v - c_i$ , and hence  $\bar{\delta} = v$ ) the log ergodic odds of state  $z$  versus state  $z'$  is proportional to the difference in their private feasibilities.*

Proposition 8 reveals that the long-run probability of a state depends upon private incentives.<sup>26</sup> The team-failure state in  $Z_0$  has a potential of zero. If  $v < \sum_{i=1}^n c_i$  then  $\psi(z) < 0$  for

<sup>25</sup>Blume (1997) observed that detailed-balance conditions are satisfied for play evolving by a log-linear choice rule (that is, logit quantal response) if and only if the game admits an exact potential function.

<sup>26</sup>Proposition 8 applies to a more general class of public-good provision games (Myatt and Wallace, 2006). Allowing each player's action  $z_i$  to be chosen from a finite set at a private cost of  $c_i(z_i)$  and writing  $G(z)$  for the production of a public good, set a player's payoff to  $u_i(z) \equiv G(z) - c_i(z_i)$ . This general game has an exact potential  $\psi(z) = G(z) - \sum_{i=1}^n c_i(z_i)$ , and Proposition 8 continues to hold.

all  $z \in Z_m$ , and so a team's success is less likely than team failure for any  $\varepsilon > 0$ . This means that in the long run the failure of a collective action stems not from the coordination problem emerging from the teamwork dilemma, but rather from the classic problem described by Olson (1968): players fail to internalise the positive externalities of participation.

Attention now returns to the bad-apple effects that were central to Propositions 1–7. Recall that a bad apple can either be (i) the weak link within the most enthusiastic team; or (ii) a player from outside the most enthusiastic team who destabilises the team by supplanting a more reliable team member. For the symmetric-valuation game considered here, a player's birth cost  $\beta_i$  and death cost  $\delta_i$  are both determined by the contribution cost  $c_i$ . A bad apple will tend to be a player with an intermediate value of  $c_i$ : low enough so that the player will volunteer now and again, but high enough to ensure that the player tends to abandon a successfully operating team with relatively high probability. The bad-apple effect is present for non-vanishing noise if the presence of such a player tends to hurt welfare.

Welfare is the utilitarian sum of payoffs:  $w(z) \equiv \sum_{i=1}^n u_i(z) = nv \times \mathcal{I}[|z| \geq m] - \sum_{i=1}^n z_i c_i$ . In the long run,  $z$  is played with probability  $p_z$ , and hence (long run) expected welfare is

$$W \equiv \sum_{z \in Z} [p_z \times w(z)] = \sum_{z \in Z} \left[ \frac{\exp[\psi(z)/\varepsilon]}{\sum_{z' \in Z} \exp[\psi(z')/\varepsilon]} \times w(z) \right].$$

$W$  reflects the tension between private and social interests: whereas welfare in state  $z$  is  $w(z)$ , its ergodic probability is determined by the private feasibility  $\psi(z)$ , which neglects externalities of  $(n-1)v \times \mathcal{I}[|z| \geq m]$ . Proposition 9 confirms that bad apples hurt welfare.

**Proposition 9.** *Social welfare is a strictly quasi-convex function of each player's private cost of contribution: starting from zero,  $W$  is first decreasing in  $c_i$  and then increasing in  $c_i$ .*

To prove this claim informally, note that a small increase in  $c_i$  has two effects: (i) it reduces welfare  $w(z)$  for all states in which  $i$  contributes; and (ii) it makes those states less likely. The latter effect may help if player  $i$  is a relatively inefficient contributor. Hence, if (ii) dominates (i), then expected welfare will increase. If this is so, consider a further increase in  $c_i$ . The negative effect (i) is felt less severely, since  $i$  contributes less frequently. Effect (ii), however, is enhanced. It is now more desirable to stop  $i$  participating, because of the higher contribution cost. Summarising: once (ii) dominates (i), then this dominance continues for higher values of  $c_i$ . This ensures that  $W$  is a quasi-convex function of  $c_i$ .<sup>27</sup>

Since  $W$  is in essence a “U-shaped” function of  $c_i$ , players with intermediate contribution costs are bad for welfare; it is (socially) preferable for costs to be low (so that a player is an efficient contributor) or high (so that the player rarely volunteers, and hence does not destabilise successful teams of enthusiasts). For the latter case, allowing  $c_i \rightarrow \infty$  is equivalent to excluding player  $i$  from the contributor pool. A straightforward corollary of Proposition 9

<sup>27</sup>The same logic was used by Myatt and Wallace (2006, Theorem 1). They provided an evolutionary justification for the use of thresholds in collective-action problems.

is that when a player has a sufficiently high contribution cost (so that  $c_i \geq c_i^*$ ) expected welfare would be enhanced by preventing player from  $i$  volunteering.

This argument suggests a rationale for limiting the size of the contributor pool in scenarios such as open-source software development. One of way of doing this is to erect barriers to participation. For instance, if the inner-workings of an open-source project are particularly technical, than relatively low-ability programmers will find it far too costly to participate, whereas high-ability programmers may be hardly affected; this will serve to exclude the unwanted bad apples. All well and good; but who should be excluded?

To answer, write  $W_i^+$  for the welfare from a process where player  $i$  is forced to participate, and  $W_i^-$  for welfare when  $i$  is excluded. The proof of Proposition 10 (Appendix C) confirms that it is optimal to admit  $i$  to the contributor pool (so that  $W > W_i^-$ ) if and only if  $W_i^+ > W_i^-$ . That is, player  $i$  should be allowed to contribute voluntarily if and only if a social planner would ideally coerce player  $i$  into participating. Clearly, such coercion is more desirable when player  $i$ 's contribution cost is low. More generally, if a social planner could choose the pool of contributors (whilst ensuring that the benefits from any provision continue to be enjoyed by all) then the contributor pool should be limited to the more efficient players.

**Proposition 10.** *Consider a symmetric-valuation public-good provision game evolving via logit quantal response. Welfare is enhanced by excluding player  $i$  from the contributor pool if and only if  $c_i$  is sufficiently large. The welfare-maximising contributor pool is either (i) empty, so that no player may volunteer; or (ii) consists of players  $\{1, \dots, n^*\}$  for some  $n^* \geq m$ . If the collective action is privately feasible, then  $n^* = m$  for  $\varepsilon$  sufficiently small.*

Exploring the final claim, when the collective action is privately feasible  $z^\dagger$  is stochastically stable and so  $W \rightarrow w(z^\dagger)$  as noise vanishes so long as the members of  $z^\dagger$  are allowed to participate. Now,  $w(z^\dagger) - W_i^+ \geq \beta_{m+1} - \beta_m > 0$  for all  $i > m$ , and hence (for small  $\varepsilon$ ) it is optimal to exclude all players  $i > m$ : they can only harm the efficiency of provision.

This section concludes by returning to the classic message of Olson (1968). His focus was the relationship between the likelihood of failure of a collective action and the size of the group involved, where his notion of group size corresponded to the number  $n$  who consume the public good. He assumed that, for a fixed private valuation of a public good, the cost of provision increases with  $n$ , and so he predicted that large groups would fail to provide public goods. Here, this comparative-static exercise might correspond to an increase in  $\sum_{i=1}^m c_i$  following an increase in  $n$ . This reduces private feasibility, and hinders team success.

There is, however, an alternative interpretation of group size: the depth of the pool of potential contributors, and hence the number who may supply the public good. Here, the results of this paper provide new insights. Whereas it may be optimal for the contributor pool to exceed the critical mass required for success (that is, it may be that  $n^* > m$  for larger  $\varepsilon$ ) nevertheless the bad-apple effect serves to limit the effectiveness of large player sets. A collective action can succeed when it relies on a select group of enthusiasts.

## APPENDIX A. THE ERGODIC DISTRIBUTION FOR VANISHING NOISE

Long-run play depends upon the rates at which transition probabilities vanish as  $\varepsilon \rightarrow 0$ . Such a rate is the “exponential cost”  $\mathcal{E}$  of a probability (Myatt and Wallace, 2003).  $\mathcal{E} \in \mathbb{R}_+ \cup \{\infty\}$  is defined for a continuous function  $p(\varepsilon)$  if either  $p(\varepsilon) = 0$  for all  $\varepsilon > 0$ , in which case  $\mathcal{E} = \infty$ , or if the limit  $\mathcal{E} = -\lim_{\varepsilon \downarrow 0} \varepsilon \log p(\varepsilon)$  exists. This property is denoted  $p(\varepsilon) = \tilde{o}(\mathcal{E})$  or  $\mathcal{E}(p(\cdot)) = \mathcal{E}$ , and means that  $p(\varepsilon)$  behaves as  $\exp(-\mathcal{E}/\varepsilon)$  does as  $\varepsilon \rightarrow 0$ . For a set  $\{p_l(\varepsilon)\}$  with exponential costs  $\{\mathcal{E}_l\}$ ,

$$\prod \tilde{o}(\mathcal{E}_l) = \tilde{o}\left(\sum \mathcal{E}_l\right), \quad \sum \tilde{o}(\mathcal{E}) = \tilde{o}(\min \mathcal{E}_l), \quad a \times \tilde{o}(\mathcal{E}) = \tilde{o}(\mathcal{E}), \quad \text{and } \mathcal{E}_l > \mathcal{E}_{l'} \Rightarrow \lim_{\varepsilon \rightarrow 0} \frac{\tilde{o}(\mathcal{E}_l)}{\tilde{o}(\mathcal{E}_{l'})} = 0. \quad (8)$$

Given the exponential-cost definition, the birth cost of a contribution against the flow of play is  $\beta_i \equiv \mathcal{E}(b_i)$ . Similarly, the death cost is  $\delta_i \equiv \mathcal{E}(d_i)$ , and moreover  $\mathcal{E}(1 - b_i) = \mathcal{E}(1 - d_i) = 0$ . The specification (1) in the text yields well-defined birth and death costs, and when play of a threshold public-good provision game evolves via logit quantal-response, these satisfy  $\beta_i = c_i$  and  $\delta_i = v_i - c_i$ . Lemma 1 confirms that birth and death costs also exist for probit quantal responses.

**Lemma 1.** *If play of the collective-action game evolves by probit quantal response, so that player  $i$  chooses to contribute if and only if  $\Delta \tilde{u}_i(z) > 0$  where  $\Delta \tilde{u}_i(z) \sim N(\Delta u_i(z), \varepsilon \times \sigma_i^2(z))$ , then*

$$\mathcal{E}(\Pr[\Delta \tilde{u}_i(z) > 0]) \equiv -\lim_{\varepsilon \rightarrow 0} \varepsilon \log \Pr[\Delta \tilde{u}_i(z) > 0] = \frac{[\Delta u_i(z)]^2}{2 \times \sigma_i^2(z)}. \quad (9)$$

When play of threshold public-good provision game evolves via probit quantal-response where (3) and (4) hold, then  $\mathcal{E}(b_i) = \beta_i$  and  $\mathcal{E}(d_i) = \delta_i$  where  $\beta_i = c_i^2/(2\xi_i^2)$  and  $\delta_i = (v_i - c_i)^2/(2\gamma_i^2)$ .

*Proof.* If  $\Delta u_i(z) > 0$  then  $\Pr[\Delta \tilde{u}_i(z) > 0] \rightarrow 1$  as  $\varepsilon \rightarrow 0$ , and so  $\mathcal{E}(\Pr[\Delta \tilde{u}_i(z) > 0]) = 0$ . If  $\Delta u_i(z) < 0$  then write  $\mathcal{E}$  for the right-hand side of (9).  $\Pr[\Delta \tilde{u}_i(z) > 0] = 1 - \Phi(x)$  where  $x = \sqrt{2\varepsilon/\varepsilon}$  and  $\Phi(\cdot)$  is the distribution of the standard normal. From a change of variable from  $\varepsilon$  to  $x$ ,

$$-\lim_{\varepsilon \rightarrow 0} [\varepsilon \times \log \Pr[\Delta \tilde{u}_i(z) > 0]] = \mathcal{E} \times \lim_{x \rightarrow \infty} \left[ -\frac{2 \log[1 - \Phi(x)]}{x^2} \right] = \mathcal{E} \times \lim_{x \rightarrow \infty} \left[ \frac{\phi(x)/[1 - \Phi(x)]}{x} \right] = \mathcal{E},$$

where  $\phi(\cdot)$  is the density of the standard normal. The penultimate equality follows from an application of l'Hôpital's rule as  $x \rightarrow \infty$ , and the final equality follows from the asymptotic linearity of the hazard rate of the normal distribution. The remaining claims of the lemma follow from substitution of the expressions for  $\Delta u_i(z)$  and  $\sigma_i^2(z)$  from (3) and (4) in the main text.  $\square$

Lemma 1 verifies (5) in the text, so that birth and death costs are defined for the probit specification. They are also defined for a wider class of models. Following Footnote 11, suppose that the noise in  $\Delta \tilde{u}_i(z)$  is drawn from the generalised error distribution (equivalently, the exponential power distribution). This has a density  $f(x) \propto \exp(-|x|^\nu)$ , where  $\nu$  is a tail thickness parameter; the normal is obtained for  $\nu = 2$ . Exponential costs then take the form  $\mathcal{E} \propto [\Delta u_i(z)]^\nu$ .

If birth and death costs are defined, (8) ensures that the exponential costs of transition probabilities are defined. Writing  $\mathcal{E}_{zz'} \equiv \mathcal{E}(\Pr[z \rightarrow z'])$ , an application of (8) yields the following lemma.

**Lemma 2.** *Suppose  $z' \neq z$ . If there is no  $i$  s.t.  $z' = z^{i+}$  or  $z' = z^{i-}$  then  $\mathcal{E}_{zz'} = \infty$ . Else,*

$$z' = z^{i+} \Rightarrow \mathcal{E}_{zz'} = \begin{cases} \beta_i & z \notin Z_{m-1}, \\ 0 & z \in Z_{m-1} \end{cases} \quad \text{and} \quad z' = z^{i-} \Rightarrow \mathcal{E}_{zz'} = \begin{cases} \delta_i & z \in Z_m, \\ 0 & z \notin Z_m. \end{cases}$$

For  $\varepsilon > 0$ , there is a unique ergodic distribution  $p = \{p_z\}_{z \in Z}$ . A graph-theoretic technique will be used to characterise  $p$  as  $\varepsilon \rightarrow 0$ . A “tree rooted at  $z$ ” is a directed graph (a subset  $h \subseteq Z \times Z$ ) such that each node  $z' \neq z$  has a unique successor. All sequences of edges lead to  $z$ , which has no successor. The set of trees rooted at  $z$  is  $H_z$ . From Freidlin and Wentzell (1998):

**Lemma 3.**  $p$  satisfies  $p_z = q_z / \sum_{z' \in Z} q_{z'}$ , where  $q_z = \sum_{h \in H_z} \prod_{(s,s') \in h} \Pr[s \rightarrow s']$ .

The relative likelihood of  $z$  and  $z'$  may be assessed via  $q_z/q_{z'}$ . Unfortunately the expression in Lemma 3 may be complicated in general. However, as  $\varepsilon \rightarrow 0$  only certain trees matter, greatly simplifying calculations. Abusing notation, write  $\mathcal{E}_h \equiv \sum_{(z,z') \in h} \mathcal{E}_{zz'}$  for the exponential cost of the product of the transition probabilities taken from the branches of the tree. Applying (8),

$$\mathcal{E}(q_z) = \mathcal{E} \left( \sum_{h \in H_z} \prod_{(s,s') \in h} \Pr[s \rightarrow s'] \right) = \min_{h \in H_z} \mathcal{E} \left( \prod_{(s,s') \in h} \Pr[s \rightarrow s'] \right) = \min_{h \in H_z} \mathcal{E}_h.$$

From (8),  $\mathcal{E}(q_z) < \mathcal{E}(q_{z'}) \Rightarrow \lim_{\varepsilon \rightarrow 0} [q_{z'}/q_z] = 0$ , so a tree with a root at  $z$  that has a lower exponential cost than any tree rooted at  $z'$  has infinitely more weight in the limit. Thus the states with minimum-exponential-cost rooted trees are “selected” as  $\varepsilon \rightarrow 0$ ; they are stochastically stable.

**Lemma 4.** States in  $Z^\dagger$  attract all probability in the limit:  $\lim_{\varepsilon \rightarrow 0} \sum_{z \in Z^\dagger} p_z = 1$ , where

$$Z^\dagger = \left\{ z \in Z : \min_{h \in H_z} \{\mathcal{E}_h\} \leq \min_{z' \in Z} \min_{h' \in H_{z'}} \{\mathcal{E}_{h'}\} \right\}.$$

A further abuse of notation is this:  $\mathcal{E}(z)$  is the exponential cost of the least-cost tree rooted at  $z$ . So, if  $\mathcal{E}(z) < \mathcal{E}(z')$  for all  $z' \neq z$ , then  $z$  is selected. Recall that, without loss, birth costs are ordered  $\beta_1 < \dots < \beta_n$ . The “most enthusiastic team” is  $z^\dagger \equiv \{z \in Z_m : z_i = 1 \Leftrightarrow i \leq m\}$ . Define  $\kappa \equiv \arg \min_{i < m} [\delta_i]$  and  $\iota \equiv \arg \min_{i \leq m} [\delta_i]$ ; these are the players with the smallest death costs amongst the first  $m-1$  and  $m$  players. Let  $\mu \equiv \arg \max_{i > m} [\delta_i]$ , the player with the largest death cost outside the most enthusiastic team. Let  $Z^\diamond = \{z \in Z_m : z_i = 1 \forall i < m \text{ and for some } i > m\}$ . The team with the  $m-1$  most enthusiastic players and some other player  $j > m$  will be labelled  $z^j \in Z^\diamond$ . Finally,  $Z^\mu = \{z^j \in Z^\diamond : \delta_j > \min[\delta_\kappa, \beta_m]\}$ , teams with the most enthusiastic  $m-1$  players and some other relatively high death-cost player.

**Theorem 1.** If  $\delta_\iota \geq \min[\delta_\mu, \beta_\iota]$  and

$$\sum_{i=1}^{m-1} \beta_i < \min[\delta^\dagger, \delta_\iota] \quad \text{where} \quad \delta^\dagger \equiv \min_{j > m} [\beta_j + \delta_j - \min[\beta_m, \delta_j]],$$

then  $\mathcal{E}(z^\dagger) < \mathcal{E}(z)$  for all  $z \neq z^\dagger$ ; the most enthusiastic team succeeds. If  $\delta_\iota < \min[\delta_\mu, \beta_\iota]$  and  $\iota = m$ ,

$$\sum_{i=1}^{m-1} \beta_i < \delta_\mu < \min[\delta_\kappa, \beta_m] \quad \Rightarrow \quad \mathcal{E}(z^\mu) < \mathcal{E}(z) \quad \text{for all } z \neq z^\mu, \text{ and}$$

$$\sum_{i=1}^{m-1} \beta_i < \min[\delta_\kappa, \beta_m] < \delta_\mu \quad \Rightarrow \quad \mathcal{E}(z^j) = \mathcal{E}(z^k) < \mathcal{E}(z) \quad \text{for all } z^j, z^k \in Z^\mu \text{ and } z \notin Z^\mu;$$

so teamwork succeeds but without player  $m$ . Otherwise  $\mathcal{E}(z^0) \leq \mathcal{E}(z)$  for all  $z \neq z^0$ ; teamwork fails.

The proof compares the least exponential-cost rooted trees at different states. Before proceeding, Lemmas 5–8 reveal the least-cost rooted trees for the states of interest. Each includes the term:

$$A = \sum_{z \in Z_m} \min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i].$$



Inspection reveals that this is the sum of the least exponential-cost exits from each state in  $Z_m$ .

**Lemma 5.** *The least exponential-cost tree rooted at state  $z^\dagger$  satisfies*

$$\mathcal{E}(z^\dagger) = A - \min[\delta_\iota, \beta_{m+1}] + \sum_{i=1}^{m-1} \beta_i.$$

*Proof.* Consider a tree rooted at  $z^\dagger$ . There is no exit from  $z^\dagger$ , but exits are required for all  $z \in Z_m$  where  $z \neq z^\dagger$ . Total costs of exits from  $Z_m$  are therefore at least  $A - \min_{i \in N} [z_i^\dagger \delta_i + (1 - z_i^\dagger) \beta_i]$ . The tree includes a path of transitions from  $Z_0$  into  $Z_m$ . Such a path involves a transition from  $Z_0$  to  $Z_1$ , from  $Z_1$  to  $Z_2$ , and so on. Each such transition includes an additional birth cost. The least costly way to reach  $Z_{m-1}$  is to acquire the  $m-1$  players with the least birth costs; this has exponential cost of at least  $\sum_{i=1}^{m-1} \beta_i$ . A final step into  $Z_m$  (adding player  $m$ ) may be taken at zero cost, and gives a path that moves as quickly (and cheaply) as possible from  $Z_0$  to  $z^\dagger \in Z_m$ . Hence

$$\mathcal{E}(z^\dagger) \geq A - \min_{i \in N} [z_i^\dagger \delta_i + (1 - z_i^\dagger) \beta_i] + \sum_{i=1}^{m-1} \beta_i. \quad (10)$$

This expression provides a lower bound to the exponential cost of any tree rooted at  $z^\dagger$ . It incorporates transitions out of all states in  $Z_m$ , as well as the path from  $Z_0$ . Notice that the cheapest birth out of  $z^\dagger$  comes at a cost of  $\beta_{m+1}$ , since  $z_i^\dagger = 1$  for all  $i \leq m$  and  $z_i^\dagger = 0$  for all  $i > m$ . Similarly, the cheapest death out of  $z^\dagger$  comes at a cost of  $\delta_\iota$ . Hence  $\min_{i \in N} [z_i^\dagger \delta_i + (1 - z_i^\dagger) \beta_i]$  in (10) can be replaced by  $\min[\delta_\iota, \beta_{m+1}]$ , yielding  $\mathcal{E}(z^\dagger) \geq A - \min[\delta_\iota, \beta_{m+1}] + \sum_{i=1}^{m-1} \beta_i$ .

It is now shown that transitions from remaining states may be constructed at no additional exponential cost, so that a tree rooted at  $z^\dagger$  is obtained that attains the lower bound in (10). To do this, take a state  $z \in Z_m$  where  $z \neq z^\dagger$  such that  $\min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i] = \delta_j$  for some  $z_j = 1$ . Hence the least-cost exit from  $z$  is to move to  $z' = z^{j-} \in Z_{m-1}$ . One possibility is that  $z'$  is encountered on the upward path from  $Z_0$  to  $z^\dagger$ . In this case, there is a path from  $z$  into  $z^\dagger$ . A second possibility is that a zero-cost step up might be taken directly into  $z^\dagger$ . If neither opportunity is available, take a zero-cost step down into  $z'' \in Z_{m-2}$ . Now,  $z''$  may be on the upward path from  $Z_0$  to  $z^\dagger$ . If not, then another zero exponential-cost step may be taken down into  $Z_{m-3}$ . This sequence continues until the graph hits either the upward path or  $Z_0$ . In either case, following the initial death of  $j$  from  $z \in Z_m$ , there is a path with zero additional exponential cost leading into  $z^\dagger$ .

Next, take a state  $z \in Z_m$  where  $z \neq z^\dagger$  such that  $\min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i] = \beta_j \neq \delta_k$ , for some  $z_j = 0$ , and for all  $z_k = 1$ . Hence the least-cost exit from  $z$  is to move up to  $z' = z^{j+} \in Z$ . From  $z'$ , construct a zero exponential-cost transition back down into  $Z_m$ . To do this, pick  $k = \arg \max_{i \in N} [\beta_i z_i]$ , and remove player  $k$ . Now  $z \neq z^\dagger$ , so  $\beta_k > \beta_j$ .<sup>28</sup> Do this for all states  $z \in Z_m$  where the least-cost exit is upward. Define  $B(z) = \sum_{i \in N} \beta_i z_i$ . Notice that, as the process moves up from state  $z \in Z_m$ , and back down again,  $B(z)$  is decreasing. Following this path, eventually state  $z^\dagger = \arg \min_{z \in Z_m} B(z)$  must be reached. Consequently there are no cycles, and there is a path from every state in  $Z_m$  to  $z^\dagger$ . Finally, consider any states above layer  $Z_m$  that are unconnected. At zero cost, remove a contributor to transit to the layer below. Similarly, for states below  $Z_m$  that are unconnected, do the same. This is a rooted tree with an exponential cost that attains the lower bound.  $\square$

<sup>28</sup>This is where the genericity assumption (see Footnote 13) that  $\beta_i \neq \beta_j$  for all  $i$  and  $j$  comes in. If there were ties then the tied players could be re-ordered by shoe size and the proof would continue.

**Lemma 6.** For all  $z \in Z_m$ , the exponential cost of a tree rooted at  $z$  is bounded:

$$\mathcal{E}(z) \geq A - \min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i] + \sum_{i=1}^{m-1} \beta_i.$$

*Proof.* Follow the argument leading to (10) in the proof of Lemma 5.  $\square$

**Lemma 7.** Define  $z^0 = z \in Z_0$ . If  $\delta_\iota \leq \beta_{m+1}$  then  $\mathcal{E}(z^0) = A$ . Alternatively, if  $\delta_\iota > \beta_{m+1}$  then

$$\mathcal{E}(z^0) = A - \beta_{m+1} + \min[\delta^\dagger, \delta_\iota] \quad \text{where} \quad \delta^\dagger \equiv \min_{j > m} [\beta_j + \delta_j - \min[\beta_m, \delta_j]].$$

*Proof.* Consider a tree rooted at  $z^0$ . There must be a transition out of every state in  $Z_m$ , and hence such a tree has an exponential cost of at least  $A$ . For states  $z \in Z_m$  where  $z \neq z^\dagger$  follow the procedure described in the proof of Lemma 5 to construct a sequence of zero-cost transitions that connect each such state to  $z^\dagger$ . For any unconnected states above or below  $Z_m$  remove a player at zero cost from the active team to yield a transition to the layer below. Finally, consider state  $z^\dagger$ . If  $\delta_\iota \leq \beta_{m+1}$  then a least exponential-cost exit is to move down to layer  $Z_{m-1}$ . This yields a tree rooted at  $z^0$  with exponential cost  $A$ . The remainder of the proof is for the case  $\delta_\iota > \beta_{m+1}$ .

Step (i). Construct a benchmark graph in the following manner. For each state in  $Z_m$ , take the least-cost exit. For all other states except  $z^0$ , take any zero-cost exit—there will always be at least one such transition available. The graph so connected has an exponential cost of  $A$ , by construction.

Step (ii). Consider a least exponential-cost tree rooted at  $z^0$ . (This will be compared to the benchmark graph from step (i).) Without loss of generality, build transitions from all states in  $Z_{m-1}$  that point downwards, at zero exponential cost, into  $Z_{m-2}$ .

Step (iii). With the tree described in step (ii) in mind, begin with state  $z^\dagger$ . Follow the sequence of directed edges in the tree until the first node in  $Z_{m-1}$  is reached. The final edge in this sequence, which terminates at a state in  $Z_{m-1}$ , must originate from some state in  $Z_m$ . Call this originating state  $z^\diamond$ . From state  $z^\diamond$ , some player  $j$  satisfying  $z_j^\diamond = 1$  must die at a cost of  $\delta_j$  to move into layer  $Z_{m-1}$ . (In preparation for step (v) below, note that it might well be that  $z^\diamond = z^\dagger$ .)

Step (iv). Take the set of states  $z \in Z_m$  that are *not* encountered on the path from  $z^\dagger$  to  $z^\diamond$ . Take the least-cost exit from each of these. Add zero-cost links from other states not in  $Z_m$  to construct transitions that lead into the aforementioned path. (This follows the procedure described in the proof of Lemma 5, and can be completed at zero exponential cost.) Any difference in the exponential cost of the tree, relative to the benchmark graph, must arise on the path from  $z^\dagger$  to  $z^\diamond$ .

Step (v). Suppose that  $z^\dagger = z^\diamond$ . If this is the least exponential-cost rooted tree, then the dying player  $j$  must be  $j = \iota$ , since this player faces the lowest death cost  $\delta_\iota$  among the team members at  $z^\dagger$ . Relative to the benchmark, there is no exit upward. The exponential cost of this tree is

$$\mathcal{E}(z^0) = A + \delta_\iota - \beta_{m+1}. \quad (11)$$

Step (vi). Suppose instead that  $z^\dagger \neq z^\diamond$ . Write  $\tilde{N} = \{i : z_i^\diamond = 1 \text{ and } z_i^\dagger = 0\}$  for the contributors acquired along the path from  $z^\dagger$  to  $z^\diamond$ . If  $i \in \tilde{N}$  then player  $i$  must have been acquired at a cost of  $\beta_i \geq \beta_{m+1}$ . If this player was acquired following an exit from  $z' \in Z_m$ , then relative to the baseline graph the net exponential cost increase must be at least  $\beta_i - \beta_{m+1}$ . In particular, the acquisition of the player of interest (the player  $j$  who dies to enable the final escape from state  $z^\diamond$ )

came at a positive net exponential cost. For  $i \neq j$ , these net costs could be saved by eliminating such acquisitions along the path. Hence, if the tree in question is a minimum-cost rooted tree, then either  $\tilde{N} = \{j\}$  or there is an equally costly tree with  $\tilde{N} = \{j\}$ . Hence the path from  $z^\dagger$  consists of two steps: the acquisition of player  $j \geq m+1$ , followed by the loss of some other player  $i \leq m$  from the original team. In fact, player  $i = m$  should be removed to ensure that the tree has minimum exponential cost, as  $m$  is the player who ensures that the cheapest possible exit-birth cost (to be removed from  $z^\diamond$ ) is as high as possible. Relative to the baseline graph (with exponential cost  $A$ ) there are additions of  $\beta_j$  and  $\delta_j$ , and deductions of  $\beta_{m+1}$  and  $\min[\delta_j, \beta_m]$ . (This last point is because, from the viewpoint of state  $z^\diamond$ , player  $m$  is the player outside the team with the least exit-birth cost.) Hence, the exponential cost of this rooted tree is  $\mathcal{E}(z^0) = A + \beta_j + \delta_j - \beta_{m+1} - \min[\beta_m, \delta_j]$ . Such a tree can be constructed by choosing *any*  $j > m$  to be added to the original team  $z^\dagger$ . Hence, if the tree in question is a minimum exponential-cost tree,  $j > m$  must minimize  $\beta_j + \delta_j - \min[\beta_m, \delta_j]$ .

Step (vii). Add back in the term  $A - \beta_{m+1}$ , to give  $\mathcal{E}(z^0) = A - \beta_{m+1} + \min_{j>m} [\beta_j + \delta_j - \min[\beta_m, \delta_j]]$ . Now  $z^\diamond \neq z^\dagger$  if and only if this term is smaller than the one given in (11) since only the least exponential-cost tree rooted at  $z^0$  is of interest. Hence that cost is

$$\mathcal{E}(z^0) = A - \beta_{m+1} + \min \left[ \min_{j>m} [\beta_j + \delta_j - \min[\beta_m, \delta_j]], \delta_\iota \right],$$

which (given the definition of  $\delta^\dagger$ ) is the required expression in the statement of the lemma.  $\square$

**Lemma 8.** *If  $\delta_\iota < \min[\delta_\mu, \beta_\iota]$  and  $\iota = m$ , then for all  $z^j \in Z^\diamond$ ,*

$$\mathcal{E}(z^j) = A - \min[\delta_j, \delta_\kappa, \beta_m] + \sum_{i=1}^{m-1} \beta_i.$$

*Proof.* By Lemma 6,  $\mathcal{E}(z^j)$  is at least  $A - \min_{i \in N} [z_i^j \delta_i + (1 - z_i^j) \beta_i] + \sum_{i=1}^{m-1} \beta_i$ . The team  $z^j$  contains  $j$ , therefore an element of the second term is  $\delta_j$ . It does not contain  $m$ , but does contain each  $i < m$ . Therefore the lowest birth cost such that  $z_i^j = 0$  is  $\beta_m$ . Finally, the lowest death cost may be  $\delta_j$  or it may be the lowest death cost of the first  $m-1$  players,  $\delta_\kappa$ . Thus

$$\mathcal{E}(z^j) \geq A - \min[\delta_j, \delta_\kappa, \beta_m] + \sum_{i=1}^{m-1} \beta_i.$$

To prove the lemma it need only be shown that there is a rooted tree that attains this bound. Consider the least exponential-cost rooted tree at  $z^\dagger$  constructed in Lemma 5. There must have been a path out of  $z^j$ , and given the construction in Lemma 5, this came at cost  $\min[\delta_j, \delta_\kappa, \beta_m]$ . Remove this branch, reducing the weight of the tree by the associated cost. A new branch is required out of  $z^\dagger$  leading to  $z^j$ . The cheapest such branch has cost  $\min[\beta_{m+1}, \delta_\iota]$ . But  $\delta_\iota < \min[\max_{i>m} \delta_i, \beta_\iota]$  and  $\iota = m$ . Hence  $\delta_\iota < \beta_\iota = \beta_m < \beta_{m+1}$ . The cheapest exit from  $z^\dagger$  is to kill player  $m$  (at the cost of  $\delta_\iota$ ). Adding player  $j$  at zero cost from the resultant state in  $Z_{m-1}$  leads directly to  $z^j$ .  $\square$

*Proof of Theorem 1 for  $\delta_\iota \geq \min[\delta_\mu, \beta_\iota]$ .* A first claim is that  $\min\{\mathcal{E}(z^0), \mathcal{E}(z^\dagger)\} \leq \mathcal{E}(z)$  for all  $z \in Z$ . Clearly  $\mathcal{E}(z^0) \leq \mathcal{E}(z)$  when  $|z| < m$ . Similarly,  $\mathcal{E}(z^m) \leq \mathcal{E}(z)$  for all  $|z| > m$  and some  $z^m \in Z_m$ . It remains to show that  $\mathcal{E}(z^\dagger) \leq \mathcal{E}(z)$  for  $z \in Z_m$ . Using Lemmas 5 and 6, it is sufficient to show

$$A - \min[\delta_\iota, \beta_{m+1}] + \sum_{i=1}^{m-1} \beta_i \leq A - \min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i] + \sum_{i=1}^{m-1} \beta_i, \quad (12)$$

since the left-hand side is  $\mathcal{E}(z^\dagger)$  and the right-hand side is lower bound on  $\mathcal{E}(z)$  for  $z \in Z_m$ . Implying (12) it will be shown that  $\min_{i \in N}[z_i \delta_i + (1 - z_i) \beta_i] \leq \min[\delta_\iota, \beta_{m+1}]$ . Three sub-cases are considered.

(i). Suppose that  $\beta_{m+1} \leq \delta_\iota$ . In state  $z \neq z^\dagger$  there is some  $j \leq m$  such that  $z_j = 0$ . This means that  $\min_{i \in N}[z_i \delta_i + (1 - z_i) \beta_i] \leq \beta_j < \beta_{m+1} = \min[\delta_\iota, \beta_{m+1}]$ , and so (12) holds.

(ii). If  $\beta_{m+1} > \delta_\iota \geq \delta_\mu$ , then  $\min_{i \in N}[z_i \delta_i + (1 - z_i) \beta_i] \leq \min_{i \in N}[\delta_\iota, (1 - z_i) \beta_i] < \min[\delta_\iota, \beta_{m+1}]$ . The first inequality follows from the fact that  $z_j = 1$  for some  $j > m$  (since  $z \neq z^\dagger$ ), and  $j$  satisfies  $\delta_j \leq \delta_\mu \leq \delta_\iota$ . The second (strict) inequality follows from the fact that  $z_j = 0$  for some  $j \leq m$  (since  $z \neq z^\dagger$ ), and this player must have  $\beta_j < \beta_{m+1}$ . Hence (12) holds once more.

(iii). The remaining case is when  $\beta_{m+1} > \delta_\iota$  and  $\delta_\mu > \delta_\iota$ . Since  $\delta_\iota \geq \min[\delta_\mu, \beta_\iota]$ , it must be the case that  $\delta_\iota \geq \beta_\iota$ . If  $z_\iota = 0$  then  $\min_{i \in N}[z_i \delta_i + (1 - z_i) \beta_i] \leq \beta_\iota \leq \delta_\iota = \min[\delta_\iota, \beta_{m+1}]$ , so that (12) holds. If, on the other hand,  $z_\iota = 1$ , then  $\min_{i \in N}[z_i \delta_i + (1 - z_i) \beta_i] \leq \delta_\iota$ , and again (12) holds.

Therefore  $\mathcal{E}(z^\dagger) \leq \mathcal{E}(z)$  for all  $z \in Z_m$ . The second step (using Lemmas 5 and 7) is to show that

$$\mathcal{E}(z^\dagger) < \mathcal{E}(z^0) \iff \sum_{i=1}^{m-1} \beta_i < \min[\delta^\dagger, \delta_\iota] \quad \text{where} \quad \delta^\dagger = \min_{j>m} [\beta_j + \delta_j - \min[\beta_m, \delta_j]]. \quad (13)$$

(i). First suppose that  $\delta_\iota \leq \beta_{m+1}$ . By Lemma 7,  $\mathcal{E}(z^0) = A$ . By Lemma 5,  $\mathcal{E}(z^\dagger) = A - \min[\delta_\iota, \beta_{m+1}] + \sum_{i=1}^{m-1} \beta_i = A - \delta_\iota + \sum_{i=1}^{m-1} \beta_i$ . Hence  $\mathcal{E}(z^\dagger) < \mathcal{E}(z^0)$  if and only if  $\sum_{i=1}^{m-1} \beta_i < \delta_\iota$ . Now notice that  $\delta^\dagger \geq \min_{j>m} [\beta_j] = \beta_{m+1}$ . Since  $\delta_\iota \leq \beta_{m+1}$ , (13) is verified.

(ii). Now consider the alternative  $\delta_\iota > \beta_{m+1}$ . From Lemma 5,  $\mathcal{E}(z^\dagger) = A - \beta_{m+1} + \sum_{i=1}^{m-1} \beta_i$ , and direct application of Lemma 7 verifies (13). This proves the first statement in Theorem 1.  $\square$

*Proof of Theorem 1 for  $\delta_\iota < \min[\delta_\mu, \beta_\iota]$ .* First note again that any state not in either  $Z_m$  or in  $Z_0$  is trivially not a candidate for selection. For this case,  $\delta_\iota < \beta_\iota < \beta_{m+1}$ . Thus, using Lemma 7,  $\mathcal{E}(z^0) = A$ . Using Lemma 6, a sufficient condition for  $z^0$  to be selected is

$$\max_{z \in Z_m} \left[ \min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i] \right] < \sum_{i=1}^{m-1} \beta_i.$$

A team in  $Z_m$  either contains  $\iota$  or not. Partition  $Z_m$  into two such sets. Then

$$z_\iota = 1 \Rightarrow \min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i] \leq \delta_\iota, \quad \text{and} \quad z_\iota = 0 \Rightarrow \min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i] \leq \beta_\iota.$$

Recall that this is the case where  $\max[\delta_\iota, \beta_\iota] = \beta_\iota$ . Therefore

$$\max_{z \in Z_m} \left[ \min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i] \right] \leq \max[\delta_\iota, \beta_\iota] = \beta_\iota.$$

So  $z^0$  is selected if  $\beta_\iota < \sum_{i=1}^{m-1} \beta_i$ . If  $\iota \neq m$  and  $m \geq 3$  this holds. If  $\iota \neq m$  and  $m = 2$ , then  $\mathcal{E}(z^0) \leq \mathcal{E}(z)$  for all  $z$  and teamwork cannot succeed. So,  $\iota = m$  is necessary for successful teamwork.

(ii). The remaining case is when  $\delta_\iota < \min[\delta_\mu, \beta_\iota]$  and  $\iota = m$ . Successful teamwork requires  $\mathcal{E}(z) < \mathcal{E}(z^0)$  for some team  $z \in Z_m$ . If  $z_i = 0$  for some  $i < m$ , then, using Lemma 6,

$$\mathcal{E}(z) \geq A - \min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i] + \sum_{i=1}^{m-1} \beta_i \geq A - \beta_{m-1} + \sum_{i=1}^{m-1} \beta_i \geq A = \mathcal{E}(z^0),$$

where the second inequality follows from  $\min_{i \in N} [z_i \delta_i + (1 - z_i) \beta_i] \leq \beta_{m-1}$  since  $z_i = 0$  for some  $i < m$ . (Notice that the final inequality is strict whenever  $m \geq 3$ , and so  $\mathcal{E}(z) > \mathcal{E}(z^0)$ .)

Hence successful teams must satisfy  $z_i = 1$  for all  $i < m$ . Such teams are  $z^\dagger$  and the members of  $Z^\circ$ . For  $z^j \in Z^\circ$ , Lemma 8 applies. Now,  $z^\dagger$  is a candidate for selection only if  $\mathcal{E}(z^\dagger) \leq \mathcal{E}(z^j)$  for all  $z^j \in Z^\circ$ . Using Lemmas 5 and 8, this occurs only if  $\delta_m \geq \min[\delta_j, \delta_\kappa, \beta_m]$  for all  $j > m$ . However,  $\delta_m < \delta_\kappa$  by definition.  $\delta_m < \beta_m$ , and there is some  $j$  such that  $\delta_j > \delta_m$ , since  $\delta_\iota < \min[\max_{i>m}[\delta_i], \beta_\iota]$  and  $\iota = m$  by construction. Therefore  $z^\dagger$  is not selected.

This leaves only  $z^0$  and teams  $z^j \in Z^\circ$  as candidates for selection. Comparing exponential costs for such states, if  $\sum_{i=1}^{m-1} \beta_i > \min[\delta_j, \delta_\kappa, \beta_m]$  for all  $j > m$  then  $z^0$  is selected; only when  $\sum_{i=1}^{m-1} \beta_i < \min[\delta_\mu, \delta_\kappa, \beta_m]$  does teamwork succeed. Which teams succeed? Two cases arise: either  $\delta_\mu < \min[\delta_\kappa, \beta_m]$  or the reverse. If the former is true, then all  $j > m$  have lower death costs than the minimum of  $\delta_\kappa$  and  $\beta_m$ . The most stable team is the one to which the player with the *highest* of these belongs. That is, by repeated application of Lemma 8,  $\delta_\mu < [\delta_\kappa, \beta_m]$  implies  $\mathcal{E}(z^\mu) < \mathcal{E}(z)$  for all  $z \in Z^\circ$  where  $z \neq z^\mu$ . On the other hand, if  $\delta_\mu > \min[\delta_\kappa, \beta_m]$ , then there are (potentially) many teams in  $Z^\circ$  with the same least exponential cost. Indeed, repeated application of Lemma 8 reveals that all teams which contain player  $j > m$  such that  $\delta_j > \min[\delta_\kappa, \beta_m]$ , and the first  $m - 1$  players have the same cost: these are precisely the teams contained in  $Z^\mu$ .  $\square$

A careful reading of the proof of Theorem 1 reveals that for generic (Footnote 13) birth and death costs and  $m \geq 3$ ,  $\mathcal{E}(z^0) < \mathcal{E}(z)$  for all  $z \neq z^0$  whenever teamwork fails, and hence  $\sum_{z \in Z_0} p_z \rightarrow 1$  as  $\varepsilon \rightarrow 0$ . This is also true when  $m = 2$  and  $\delta_\iota \geq \min[\delta_\mu, \beta_\iota]$ . The remaining loose end is when  $m = 2$  and  $\delta_\iota < \min[\delta_\mu, \beta_\iota]$ . For this case,  $\mathcal{E}(z^0) = A$ . Suppose that  $\iota = 1 \neq m$ , and consider a team in  $Z_m$  with members  $i$  and  $j$ . If  $\min[\delta_i, \delta_j] \leq \beta_1$ , then further tree surgery confirms that  $\mathcal{E}(z) = A$ . For  $m = 2$ , then, there is a generic parameter configuration for which states in  $Z_0$  and  $Z_m$  retain positive probability as  $\varepsilon \rightarrow 0$ . This special case involves only the partial failure of collective action.

## APPENDIX B. CONCORDANCE AND TEAM SUCCESS

This appendix relates the ordering of death costs and equilibrium selection. Call  $\hat{\delta} = (\hat{\delta}_i)_{i=1}^n$  a *discordant shuffle* of  $\delta = (\delta_i)_{i=1}^m$  whenever  $C$  is more concordant than  $\hat{C}$  and for each  $i$ ,  $\hat{\delta}_i = \delta_j$  for some  $j$  (and vice-versa). Write  $\hat{\kappa} \equiv \arg \min_{i < m} [\hat{\delta}_i]$ ,  $\hat{\iota} \equiv \arg \min_{i \leq m} [\hat{\delta}_i]$ , and  $\hat{\mu} \equiv \arg \max_{i > m} [\hat{\delta}_i]$ .

**Lemma 9.** *If  $\hat{\delta}$  is a discordant shuffle of  $\delta$  then  $\hat{\delta}_{\hat{\iota}} \geq \delta_\iota$ ,  $\hat{\delta}_{\hat{\kappa}} \geq \delta_\kappa$ , and  $\hat{\delta}_{\hat{\mu}} \leq \delta_\mu$ .*

*Proof.* Let  $r(i)$  be the rank of  $\delta_i$  and  $\hat{r}(i)$  be the rank of  $\hat{\delta}_i$ , so that  $\hat{\delta}_{\hat{\iota}} = \hat{\delta}_{(\hat{r}(\hat{\iota}))}$ . If  $\hat{\delta}_{\hat{\iota}} < \delta_\iota$ , then

$$C(m, \hat{r}(\hat{\iota})) = \sum_{i=1}^m \mathcal{I}[\delta_i \leq \hat{\delta}_{\hat{\iota}}] = 0 \quad \text{and} \quad \hat{C}(m, \hat{r}(\hat{\iota})) = \sum_{i=1}^m \mathcal{I}[\hat{\delta}_i \leq \hat{\delta}_{\hat{\iota}}] = 1,$$

but  $C$  is more concordant than  $\hat{C}$ , yielding a contradiction. An analogous argument proves the lemma's second inequality. Finally suppose, again to the contrary, that  $\hat{\delta}_{\hat{\mu}} > \delta_\mu$ . This means that  $\hat{\delta}_{\hat{\mu}} = \delta_i$  for some  $i \leq m$ . Suppose  $C(m, r(\mu)) = k$  (with  $k \in \{0, m - 1\}$ , since  $\hat{\delta}_{\hat{\mu}} = \delta_i > \delta_\mu$  for some  $i \leq m$ ). Given a death cost configuration  $\delta$ , there are  $k$  death costs within the first  $m$  players lower than  $\delta_\mu$ . Every player  $j > m$  has a death cost lower than  $\delta_\mu$ . Therefore, given that (at least) one of the death costs *above*  $\delta_\mu$  no longer belongs to  $j \leq m$  under the configuration  $\hat{\delta}$ ,  $\hat{C}(m, r(\mu)) = \sum_{i=1}^m \mathcal{I}[\hat{\delta}_i \leq \delta_\mu] > k$ , contradicting the fact that  $C$  is more concordant than  $\hat{C}$ .  $\square$

In a natural notation, let the exponential cost of the least-cost tree rooted at state  $z$  under a new configuration of death costs  $\hat{\delta}$  be  $\hat{\mathcal{E}}(z)$ . Proposition 4 is a consequence of the following theorem.

**Theorem 2.** If  $\hat{\delta}$  is a discordant shuffle of  $\delta$ , then  $\mathcal{E}(z) < \mathcal{E}(z^0) \Rightarrow \hat{\mathcal{E}}(z') < \hat{\mathcal{E}}(z^0)$  for some  $z, z' \in Z$ .

*Proof of Theorem 2.* Clearly, either  $z = z^\dagger$  or  $z \in Z^\mu$ . Suppose  $z = z^\dagger$  and  $\delta_\iota \leq \beta_{m+1}$ . Then  $\mathcal{E}(z^0) = A$  and  $\mathcal{E}(z^\dagger) = A + \sum_{i=1}^{m-1} \beta_i - \delta_\iota$  from Lemmas 7 and 5. Now  $\mathcal{E}(z^\dagger) < \mathcal{E}(z^0) \Rightarrow \sum_{i=1}^{m-1} \beta_i < \delta_\iota$ . By Lemma 9,  $\hat{\delta}_\iota \geq \delta_\iota > \sum_{i=1}^{m-1} \beta_i$ . If  $\hat{\delta}_\iota \leq \beta_{m+1}$  then  $\hat{\mathcal{E}}(z^\dagger) - \hat{\mathcal{E}}(z^0) = \sum_{i=1}^{m-1} \beta_i - \hat{\delta}_\iota < 0$  and the result follows. If  $\hat{\delta}_\iota > \beta_{m+1}$  then (again from Lemmas 7 and 5),

$$\hat{\mathcal{E}}(z^\dagger) - \hat{\mathcal{E}}(z^0) = \sum_{i=1}^{m-1} \beta_i - \min[\hat{\delta}^\dagger, \hat{\delta}_\iota] \leq \sum_{i=1}^{m-1} \beta_i - \beta_{m+1} \leq \sum_{i=1}^{m-1} \beta_i - \delta_\iota < 0,$$

where  $\hat{\delta}^\dagger \equiv \min_{j>m} [\beta_j + \hat{\delta}_j - \min[\beta_m, \hat{\delta}_j]] \geq \beta_{m+1}$ .

Once again, the result follows. Now suppose that  $z = z^\dagger$  and that  $\delta_\iota > \beta_{m+1}$ . Thus, by Lemma 9,  $\hat{\delta}_\iota > \beta_{m+1}$ . Now, from Lemma 5,  $\mathcal{E}(z^\dagger) = A + \sum_{i=1}^{m-1} \beta_i - \beta_{m+1}$ , and from Lemma 7,

$$\mathcal{E}(z^0) = A - \beta_{m+1} + \min[\delta^\dagger, \delta_\iota] \quad \text{where} \quad \delta^\dagger \equiv \min_{j>m} [\beta_j + \delta_j - \min[\beta_m, \delta_j]]. \quad (14)$$

Therefore  $\sum_{i=1}^{m-1} \beta_i < \min[\delta^\dagger, \delta_\iota]$ . Now  $\hat{\mathcal{E}}(z^\dagger) - \hat{\mathcal{E}}(z^0) = \sum_{i=1}^{m-1} \beta_i - \min[\hat{\delta}^\dagger, \hat{\delta}_\iota]$  and so it is sufficient to show that  $\min[\delta^\dagger, \delta_\iota] \leq \min[\hat{\delta}^\dagger, \hat{\delta}_\iota]$ . If  $\hat{\delta}_\iota \leq \hat{\delta}^\dagger$  then  $\min[\hat{\delta}^\dagger, \hat{\delta}_\iota] \geq \hat{\delta}_\iota \geq \delta_\iota \geq \min[\delta^\dagger, \delta_\iota]$ , where the second inequality is from Lemma 9, and so  $\min[\hat{\delta}^\dagger, \hat{\delta}_\iota] \geq \min[\delta^\dagger, \delta_\iota]$  as required. If  $\hat{\delta}_\iota > \hat{\delta}^\dagger$  then define  $k = \arg \min_{j>m} [\beta_j + \hat{\delta}_j - \min[\beta_m, \hat{\delta}_j]]$ . If  $\delta_k \leq \hat{\delta}_k$  then  $\beta_k + \delta_k - \min[\beta_m, \delta_k] \leq \beta_k + \hat{\delta}_k - \min[\beta_m, \hat{\delta}_k]$ , since this term is weakly increasing in  $\delta_k$ . So

$$\min[\delta^\dagger, \delta_\iota] \leq \delta_\iota \equiv \min_{j>m} [\beta_j + \delta_j - \min[\beta_m, \delta_j]] \leq \beta_k + \delta_k - \min[\beta_m, \delta_k] \leq \beta_k + \hat{\delta}_k - \min[\beta_m, \hat{\delta}_k] = \min[\hat{\delta}^\dagger, \hat{\delta}_\iota], \quad (15)$$

as required. Suppose  $\delta_k > \hat{\delta}_k$ . Now  $\hat{C}(k, \hat{r}(k)) = \sum_{i=1}^k \mathcal{I}[\hat{\delta}_i \leq \hat{\delta}_k] = 1$ . To see this note that were there a  $j > m$  such that  $\hat{\delta}_j \leq \hat{\delta}_k$  then  $k$  could not have been the minimiser in the first place. If there were a  $j \leq m$  such that  $\hat{\delta}_j \leq \hat{\delta}_k$  then certainly  $\hat{\delta}_\iota \leq \hat{\delta}_k$  and thus  $\hat{\delta}_\iota$  would be smaller than  $\beta_k + \hat{\delta}_k - \min[\beta_m, \hat{\delta}_k]$ , contrary to assumption. Thus, for  $\hat{\delta}$  to be a discordant shuffle of  $\delta$ ,  $\hat{C}(k, \hat{r}(k)) = \sum_{i=1}^k \mathcal{I}[\delta_i \leq \hat{\delta}_k] \geq 1$ . So there exists  $i \leq k$  such that  $\delta_i \leq \hat{\delta}_k$ .  $i \neq k$  since  $\delta_k > \hat{\delta}_k$  by assumption. Thus there exists  $i < k$  where  $\delta_i \leq \hat{\delta}_k$ . For such an  $i > m$ :  $\beta_i + \delta_i - \min[\beta_m, \delta_i] \leq \beta_k + \hat{\delta}_k - \min[\beta_m, \hat{\delta}_k]$ ; (15) holds with the second inequality strict. For such an  $i \leq m$ :  $\delta_\iota \leq \delta_i \leq \hat{\delta}_k < \beta_k + \hat{\delta}_k - \min[\beta_m, \hat{\delta}_k]$  and hence  $\min[\delta^\dagger, \delta_\iota] < \beta_k + \hat{\delta}_k - \min[\beta_m, \hat{\delta}_k]$  as required.

Suppose now that  $z \neq z^\dagger$  but  $z \in Z^\mu$ . Then, by Theorem 1,  $\delta_\iota < \min[\delta_\mu, \beta_\iota]$ ,  $\iota = m$ , and  $\sum_{i=1}^{m-1} \beta_i < \min[\delta_\kappa, \delta_\mu, \beta_m]$ . First, suppose that  $\hat{\iota} \neq m$ . Consider  $z' = z^\dagger$ . If  $\hat{\delta}_\iota \leq \beta_{m+1}$  then  $\hat{\mathcal{E}}(z^\dagger) - \hat{\mathcal{E}}(z^0) = \sum_{i=1}^{m-1} \beta_i - \hat{\delta}_\iota$  by Lemmas 5 and 7. Since  $\hat{\iota} < m$ :  $\hat{\delta}_\iota = \hat{\delta}_{\hat{\kappa}} \geq \delta_\kappa \geq \min[\delta_\kappa, \delta_\mu, \beta_m] > \sum_{i=1}^{m-1} \beta_i$  (the first inequality follows from Lemma 9) and so  $\hat{\mathcal{E}}(z') = \hat{\mathcal{E}}(z^\dagger) < \hat{\mathcal{E}}(z^0)$  as required. If  $\hat{\delta}_\iota > \beta_{m+1}$ , then  $\hat{\mathcal{E}}(z^\dagger) - \hat{\mathcal{E}}(z^0) = \sum_{i=1}^{m-1} \beta_i - \min[\hat{\delta}^\dagger, \hat{\delta}_\iota]$  again by Lemmas 5 and 7. The minimum in this expression is at least  $\beta_{m+1}$ , and  $\beta_{m+1} > \beta_m \geq \min[\delta_\kappa, \delta_\mu, \beta_m] > \sum_{i=1}^{m-1} \beta_i$ . Once again  $\hat{\mathcal{E}}(z') = \hat{\mathcal{E}}(z^\dagger) < \hat{\mathcal{E}}(z^0)$  as required.<sup>29</sup> The remaining case is when  $\hat{\iota} = m$ .

First, suppose that  $\hat{\delta}_\iota < \min[\beta_\iota, \hat{\delta}_{\hat{\mu}}]$ . Consider  $z' \in Z^\mu$ .  $\hat{\mathcal{E}}(z') < \hat{\mathcal{E}}(z^0)$  if  $\sum_{i=1}^{m-1} \beta_i < \min[\hat{\delta}_{\hat{\kappa}}, \hat{\delta}_{\hat{\mu}}, \beta_m]$  (by Theorem 1). Suppose this inequality breaks. Noting that  $\hat{\delta}_{\hat{\kappa}} \geq \delta_\kappa$ , this can only happen if  $\hat{\delta}_\iota < \hat{\delta}_{\hat{\mu}} \leq \sum_{i=1}^{m-1} \beta_i < \min[\beta_m, \hat{\delta}_{\hat{\kappa}}]$ . So  $\hat{\delta}_{\hat{\mu}} < \delta_\mu$ . Thus  $\delta_\mu = \hat{\delta}_i$  for some  $i \leq m$ . Then there must exist

<sup>29</sup>Since  $\iota = m$  is a requirement for  $z \in Z^\mu$  to be selected, it must therefore follow that  $z^\dagger$  is selected if  $\hat{\iota} \neq m$ . This in turn means that the precondition  $\hat{\delta}_\iota \geq \min[\beta_\iota, \hat{\delta}_{\hat{\mu}}]$  is met (see Theorem 1). Of course, this is so, since  $\delta_\kappa > \sum_{i=1}^{m-1} \beta_i \Rightarrow \delta_\kappa > \beta_\kappa = \beta_i \geq \min[\beta_i, \hat{\delta}_{\hat{\mu}}]$  and  $\hat{\delta}_\iota = \hat{\delta}_{\hat{\kappa}} \geq \delta_\kappa$ .

some  $j > m$  such that  $\hat{\delta}_j = \delta_k$  for  $k \leq m$ . If  $k < m$ , then  $\hat{\delta}_{\hat{\mu}} \geq \delta_k \geq \delta_\kappa > \sum_{i=1}^{m-1} \beta_i$ : a contradiction. If  $k = m$ , then suppose that  $i = m$  so that  $\hat{\delta}_i = \delta_\mu > \hat{\delta}_{\hat{\mu}}$ , a contradiction. Alternatively, suppose that  $\hat{\delta}_i = \delta_\mu$  some  $i < m$ . Then  $\delta_i \geq \delta_\kappa$ . If  $\hat{\delta}_j = \delta_i$  with  $j > m$ , then once again  $\hat{\delta}_{\hat{\mu}} \geq \delta_i \geq \delta_\kappa > \sum_{i=1}^{m-1} \beta_i$ . If  $\hat{\delta}_j = \delta_i$  with  $j = m$ , then  $\hat{\delta}_i = \delta_i \geq \delta_\kappa$ , but  $\hat{\delta}_{\hat{\mu}} > \hat{\delta}_i \geq \delta_\kappa > \sum_{i=1}^{m-1} \beta_i$ : a final contradiction.

Second, suppose that  $\hat{\delta}_i \geq \min[\beta_i, \hat{\delta}_{\hat{\mu}}]$ . Consider  $z' = z^\dagger$ . If  $\hat{\delta}_i > \beta_{m+1}$  then by Lemmas 5 and 7,  $\hat{\mathcal{E}}(z^\dagger) - \hat{\mathcal{E}}(z^0) = \sum_{i=1}^{m-1} \beta_i - \min[\hat{\delta}^\dagger, \hat{\delta}_i]$ . The smallest this latter minimum could be is  $\beta_{m+1}$ . But  $\beta_{m+1} > \beta_m > \sum_{i=1}^{m-1} \beta_i$ , so  $\hat{\mathcal{E}}(z') = \hat{\mathcal{E}}(z^\dagger) < \hat{\mathcal{E}}(z^0)$ , and  $z^\dagger$  is selected. If  $\hat{\delta}_i \leq \beta_{m+1}$  then, again by Lemmas 5 and 7,  $\hat{\mathcal{E}}(z^\dagger) - \hat{\mathcal{E}}(z^0) = \sum_{i=1}^{m-1} \beta_i - \hat{\delta}_i$ . Now if  $\hat{\delta}_i \geq \beta_m$  then  $\hat{\delta}_i > \sum_{i=1}^{m-1} \beta_i$  and  $z^\dagger$  is selected. If  $\hat{\delta}_i < \beta_m$  then  $\hat{\delta}_i \geq \hat{\delta}_{\hat{\mu}}$ . Now  $\hat{\delta}_{\hat{\mu}} \leq \delta_\mu$  by Lemma 9. If  $\hat{\delta}_{\hat{\mu}} = \delta_\mu$ , then  $\hat{\delta}_i \geq \delta_\mu > \sum_{i=1}^{m-1} \beta_i$  and  $z^\dagger$  is selected. If  $\hat{\delta}_{\hat{\mu}} < \delta_\mu$  then there is an  $i \leq m$  such that  $\hat{\delta}_i = \delta_\mu$ . Therefore, there is a  $k > m$  and a  $j \leq m$  such that  $\hat{\delta}_k = \delta_j$ . If  $j < m$  then  $\delta_j \geq \delta_\kappa$  and so  $\hat{\delta}_{\hat{\mu}} \geq \hat{\delta}_k = \delta_j \geq \delta_\kappa > \sum_{i=1}^{m-1} \beta_i$ . Now  $\hat{\delta}_i \geq \hat{\delta}_{\hat{\mu}}$  and  $z^\dagger$  is selected. If  $j = m$ , then either  $i = m$  or  $i < m$ . If the former,  $\hat{\delta}_i = \delta_\mu > \sum_{i=1}^{m-1} \beta_i$  and  $z^\dagger$  is selected. If the latter, then  $\hat{\delta}_i = \delta_l$  for some  $l < m$  and so  $\hat{\delta}_i \geq \delta_\kappa > \sum_{i=1}^{m-1} \beta_i$ , and hence once again  $z^\dagger$  is selected.  $\square$

## APPENDIX C. OMITTED PROOFS

*Proof of Proposition 1.* For this birth-death process, the ergodic odds of  $Z_m$  and  $Z_0$  satisfy

$$\varepsilon \log \frac{\sum_{z \in Z_m} p_z}{\sum_{z \in Z_0} p_z} = \varepsilon \log \binom{n}{m} - (m-1)\varepsilon \log \left[ \frac{1-b}{b} \right] + \varepsilon \log \left[ \frac{1-d}{d} \right] \rightarrow \delta - (m-1)\beta \quad \text{as } \varepsilon \rightarrow 0.$$

Hence if  $\delta > (m-1)\beta$  then  $[\sum_{z \in Z_m} p_z] / [\sum_{z \in Z_0} p_z] \rightarrow \infty$  as  $\varepsilon \rightarrow 0$ . Calculations reveal that states outside  $Z_0$  and  $Z_m$  are never selected, and that  $z^0$  is selected when  $\delta < (m-1)\beta$ .  $\square$

*Proof of Proposition 2.* When  $\delta_1 > \dots > \delta_n$ ,  $\iota = m$  and  $\mu = m+1$ . The first part of Theorem 1 applies since  $\delta_m \geq \min[\delta_{m+1}, \beta_m]$  automatically. Thus  $\sum_{i=1}^{m-1} \beta_i < \min[\delta^\dagger, \delta_m]$  implies  $\mathcal{E}(z^\dagger) < \mathcal{E}(z)$  for all  $z \neq z^\dagger$  and otherwise  $\mathcal{E}(z^0) < \mathcal{E}(z)$  for all  $z \neq z^0$ , as required.  $\square$

*Proof of Proposition 3.* When  $\delta_1 < \dots < \delta_n$ ,  $\iota = 1$  and  $\mu = n$ . Now  $\delta_1 < \min[\delta_n, \beta_1]$  if and only if  $\delta_1 < \beta_1$ . But  $\beta_1 > \delta_1$  certainly implies  $\sum_{i=1}^{m-1} \beta_i > \delta_1$ . Now, because  $\iota \neq m$ ,  $\mathcal{E}(z^0) < \mathcal{E}(z)$  for all  $z \neq z^0$  from Theorem 1. Consider the alternative case  $\delta_1 \geq \min[\delta_n, \beta_1]$  (so that  $\min[\delta_n, \beta_1] = \beta_1$  must hold), and note that  $\min[\delta^\dagger, \delta_\iota] = \delta_1$ . To see this latter fact, note that  $\delta^\dagger \equiv \min_{j>m} [\beta_j + \delta_j - \min[\delta_j, \beta_m]] \geq \delta_{m+1} > \delta_1$ . Thus, applying Theorem 1 again gives

$$\sum_{i=1}^{m-1} \beta_i < \delta_1 \quad \Rightarrow \quad \mathcal{E}(z^\dagger) < \mathcal{E}(z) \quad (16)$$

for all  $z \neq z^\dagger$ . Otherwise  $\mathcal{E}(z^0) < \mathcal{E}(z)$  for all  $z \neq z^0$ . If the inequality in (16) holds then clearly  $\delta_1 > \beta_1$ . So the inequality is sufficient for selection of  $z^\dagger$ ; its reverse sufficient for selection of  $z^0$ .  $\square$

*Proof of Proposition 4.* From the application of Theorems 1 and 2.  $\square$

*Proof of Propositions 5–8.* From arguments given in the main text.  $\square$

*Proof of Proposition 9.*  $w(z) = n\mathcal{I}[|z| \geq m] - \sum_{i=1}^n z_i c_i$  and  $\psi(z) = v\mathcal{I}[|z| \geq m] - \sum_{i=1}^n z_i c_i$ . So,

$$\frac{\partial w}{\partial c_i} = \frac{\partial \psi}{\partial c_i} = \begin{cases} -1 & \text{if } z_i = 1, \text{ and} \\ 0 & \text{if } z_i = 0, \end{cases}$$

for any  $i$ . Suppressing arguments,  $W$  satisfies  $W \sum_{z \in Z} \exp[\psi/\varepsilon] = \sum_{z \in Z} w \exp[\psi/\varepsilon]$ . Write  $Z_{i+} = \{z : z_i = 1\}$  for the states in which  $i$  contributes. Differentiating and collecting terms yields

$$\frac{dW}{dc_i} \sum_{z \in Z} \exp[\psi/\varepsilon] = \sum_{z \in Z_{i+}} \exp[\psi/\varepsilon] \left[ \frac{W}{\varepsilon} - \frac{w}{\varepsilon} - 1 \right]. \quad (17)$$

Differentiating a second time with respect to  $c_i$ , and cancelling common terms gives

$$\frac{d^2 W}{dc_i^2} \sum_{z \in Z} \exp[\psi/\varepsilon] = -\frac{1}{\varepsilon} \underbrace{\sum_{z \in Z_{i+}} \exp[\psi/\varepsilon] \left[ \frac{W}{\varepsilon} - \frac{w}{\varepsilon} - 1 \right]}_{\text{zero when } dW/dc_i = 0} + \frac{1}{\varepsilon} \underbrace{\sum_{z \in Z_{i+}} \exp[\psi/\varepsilon]}_{\text{strictly positive}}.$$

At a stationary point the second derivative is strictly positive and so  $W$  is strictly quasi-convex in  $c_i$ . The right-hand side of (17) is strictly positive for  $c_i$  sufficiently large and strictly negative sufficiently small (and possibly negative).  $\square$

*Proof of Proposition 10.* Welfare can be written in terms of four components:

$$W = \frac{\sum_{z \in Z} w \exp[\psi/\varepsilon]}{\sum_{z \in Z} \exp[\psi/\varepsilon]} = \frac{\sum_{z \in Z_{i+}} w \exp[\psi/\varepsilon] + \sum_{z \notin Z_{i+}} w \exp[\psi/\varepsilon]}{\sum_{z \in Z_{i+}} \exp[\psi/\varepsilon] + \sum_{z \notin Z_{i+}} \exp[\psi/\varepsilon]} = \frac{A_i^+ + A_i^-}{B_i^+ + B_i^-}.$$

Excluding  $i$  yields welfare  $W_i^- = A_i^- / B_i^-$ . Thus  $i$  should be excluded if and only if  $W < W_i^-$ . This occurs whenever  $A_i^+ / B_i^+ < A_i^- / B_i^-$ , or equivalently  $W_i^+ < W_i^-$ , where  $W_i^+$  is welfare when player  $i$  is coerced. For the second claim, note that  $\partial W_i^- / \partial c_i = 0$ , and consider  $W_i^+$ . Since  $z_i = 1$  for all  $z \in Z_{i+}$ , an increase in  $c_i$  can only lower welfare. Thus  $W_i^+$  crosses  $W_i^-$  at most once, and there exists a  $c_i^* \in [0, \infty)$  such that if  $c_i \geq c_i^*$  welfare is larger with player  $i$  excluded.

Turning to the second claim, suppose that the optimal contributor pool contains  $j$  but not  $i < j$ . Welfare must decrease with  $c_j$ , since otherwise (due to the quasi-convexity of  $W$  in  $c_j$ ) it would be optimal to exclude  $j$ . Hence welfare is increased by lowering  $c_j$  down to  $c_i$ . But this can be achieved by removing  $j$  from the pool and including  $i$ . Hence the contributor pool was sub-optimal; a contradiction. This means that the optimal contributor pool may be empty, or it may be  $\{1, \dots, n^*\}$  for some  $n^*$ . If  $m^* < m$ , then the collective action cannot succeed, and so contributions generate costs without any benefit. Hence, if the contributor pool is non-empty, then  $n^* \geq m$ .

The final claim of the proposition follows from the argument given in the text.  $\square$

## REFERENCES

- BENDOR, J., D. DIERMEIER, AND M. TING (2003): “A Behavioral Model of Turnout,” *American Political Science Review*, 97(2), 261–80.
- BERGIN, J., AND B. LIPMAN (1996): “Evolution with State-Dependent Mutations,” *Econometrica*, 64(4), 943–56.
- BLUME, L. E. (1995): “The Statistical Mechanics of Best-Response Strategy Revision,” *Games and Economic Behavior*, 11(2), 111–45.
- (1997): “Population Games,” in *The Economy as an Evolving Complex System II*, ed. by W. B. Arthur, S. N. Durlauf, and D. A. Lane. Westview Press, Boulder, CO.
- (2003): “How Noise Matters,” *Games and Economic Behavior*, 44(2), 251–71.



- BLUME, L. E., AND S. N. DURLAUF (2001): "The Interactions-Based Approach to Socioeconomic Behaviour," in *Social Dynamics*, ed. by S. N. Durlauf, and H. P. Young, chap. 2, pp. 15–44. MIT Press, Cambridge, MA.
- CHERUBINI, U., E. LUCIANO, AND W. VECCHIATO (2004): *Copula Methods in Finance*. John Wiley and Sons, Chichester, England.
- DIERMEIER, D., AND J. A. VAN MIEGHEM (2005a): "Coordination and Turnout in Large Elections," *Northwestern University*.
- (2005b): "A Stochastic Model of Consumer Boycotts," *Northwestern University*.
- ELLISON, G. (2000): "Basins of Attraction, Long Run Stochastic Stability and the Speed of Step-by-Step Evolution," *Review of Economic Studies*, 67(1), 17–45.
- FOSTER, D., AND H. P. YOUNG (1990): "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology*, 38, 219–32.
- FREIDLIN, M. I., AND A. D. WENTZELL (1998): *Random Perturbations of Dynamical Systems*. Springer-Verlag, Berlin/New York, 2nd edn.
- HANSEN, J. M. (1985): "The Political Economy of Group Membership," *American Political Science Review*, 79(1), 79–96.
- HARVEY, A. C. (1981): *The Econometric Analysis of Time Series*. Philip Allan, Oxford.
- INNES, R. (2006): "A Theory of Consumer Boycotts under Symmetric Information and Imperfect Competition," *Economic Journal*, 116, 355–81.
- JOHNSON, J. P. (2002): "Open Source Software: Private Provision of a Public Good," *Journal of Economics and Management Strategy*, 11(4), 637–662.
- KANDORI, M., G. J. MAILATH, AND R. ROB (1993): "Learning, Mutation and Long-Run Equilibria in Games," *Econometrica*, 61(1), 29–56.
- LERNER, J., AND J. TIROLE (2002): "Some Simple Economics of Open Source," *Journal of Industrial Economics*, 50(2), 197–234.
- (2005): "The Scope of Open Source Licensing," *Journal of Law, Economics, and Organization*, 21(1), 20–56.
- MARWELL, G., P. E. OLIVER, AND R. PRAHL (1988): "Social Networks and Collective Action: A Theory of the Critical Mass. III," *American Journal of Sociology*, 94(3), 502–34.
- MAYOR, G., J. SUÑER, AND J. TORRENS (2005): "Copula-Like Operations on Finite Settings," *IEEE Transactions on Fuzzy Systems*, 13(4), 468–77.
- McKELVEY, R. D., AND T. R. PALFREY (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10(1), 6–38.
- MONDERER, D., AND L. S. SHAPLEY (1996): "Potential Games," *Games and Economic Behavior*, 14(1), 124–43.
- MYATT, D. P., AND C. WALLACE (2003): "A Multinomial Probit Model of Stochastic Evolution," *Journal of Economic Theory*, 113(2), 286–301.
- (2006): "An Evolutionary Justification for Thresholds in Collective-Action Problems," *Working Paper*.
- NELSEN, R. B. (2006): *An Introduction to Copulas*. Springer-Verlag, New York, second edn.
- OLIVER, P. E., AND G. MARWELL (1988): "The Paradox of Group Size in Collective Action: A Theory of Critical Mass. II," *American Sociological Review*, 53(1), 1–8.
- OLIVER, P. E., G. MARWELL, AND R. TEIXEIRA (1985): "A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action," *American Journal of Sociology*, 91(3), 522–56.
- OLSON, M. (1968): *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, Cambridge, MA.
- PALFREY, T. R., AND H. ROSENTHAL (1984): "Participation and the Provision of Discrete Public Goods: A Strategic Analysis," *Journal of Public Economics*, 24(2), 171–93.
- (1985): "Voter Participation and Strategic Uncertainty," *American Political Science Review*, 79(1), 62–78.
- YOUNG, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61(1), 57–84.