THE REVIEW OF ECONOMIC STUDIES

Review of Economic Studies (2008) **75**, 499–527 0034-6527/08/00210499$02.00

When Does One Bad Apple Spoil the Barrel? An Evolutionary Analysis of Collective Action

DAVID P. MYATT

Department of Economics and Nuffield College, University of Oxford

and

CHRIS WALLACE

Department of Economics and Trinity College, University of Oxford

First version received June 2006; *final version accepted September* 2007 (*Eds.*)

This paper studies collective-action games in which the production of a public good requires teamwork. A leading example is a threshold game in which provision requires the voluntary participation of m out of n players. Quantal-response strategy revisions allow play to move between equilibria in which a team successfully provides, and an equilibrium in which the collective action fails. A full characterization 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 destabilizing successful teams and so offers a rationale for limiting the pool of possible contributors.

1. COLLECTIVE ACTION

The problems of collective action have long concerned social scientists. To economists, the Olson (1968) view is that a collective action involves the private production of a public good or the private exploitation of a common-pool resource; the associated externalities lead to socially suboptimal public-good provision or a tragedy of the commons. Political scientists and sociologists have also found many instances of these problems. These include voter participation (Palfrey and Rosenthal, 1985; Bendor, Diermeier and Ting, 2003; Diermeier and Van Mieghem, 2005*a*), interest groups (Oliver, Marwell and Teixeira, 1985; Marwell, Oliver and Prahl, 1988; Oliver and Marwell, 1988), industry lobbies (Hansen, 1985), consumer boycotts (Diermeier and Van Mieghem, 2005*b*; Innes, 2006), and many others.

In the context of privately provided public goods, Olson's (1968) strong free-rider hypothesis is that contributions fall with the size of the group involved. This requires the public good to be at least partially rival, so that increased consumption raises the cost of provision. This jointness of consumption relates provision incentives to the size of the consuming group. However, this view does not reveal the relationship between such incentives and the number of other contributors. This latter relationship may stem from the jointness of supply, which applies when the optimal provision decision of an individual depends upon the contributions of others. For example, if public-good production were a concave function of aggregate private contributions then a standard Cournot-contributions game would arise (Shibata, 1971; Warr, 1983; Cornes and Sandler, 1984; Bergstrom, Blume and Varian, 1986). In such a game an individual's incentive to provide falls with the contributions of others.

However, in other circumstances there is an incentive to provide only if others do so. The teamwork dilemmas that arise in such situations are precisely the focus of this paper. These dilemmas are most simply captured by an n-player "threshold" binary-action game in which m or more voluntary contributions are required for the production of a public good (Palfrey and Rosenthal, 1984). Such a game reflects the collective nature of collective action. The jointness of supply offers a possible resolution to the underprovision problem: when an entire project depends in a pivotal way on the contributions of each individual, incentives are dramatically enhanced, so long as others play their part. On the other hand, when $m > 1$ there is an equilibrium in which the team (and with it the collective project) fails. Beyond economics, threshold games of this form are also central to a large and important sociological literature. The need for a critical level of participation corresponds to the "minimal contributing set" of van de Kragt, Orbell and Dawes (1983) and Rapoport (1985).[1] The equilibrium-selection problem was (implicitly) recognized in sociological theories of "critical mass" (Oliver and Marwell, 2001) and models of crowd behaviour (Berk, 1974).

The critical feature of a threshold game is that a convexity in the public-good production function generates a coordination problem; the "n choose m" specification is the simplest way to capture this feature. The insights are substantially unchanged when attention moves to a richer class of public-good provision games. Threshold games are amenable to detailed analysis with fully asymmetric players. However, later in the paper the focus broadens out to a completely general class of public-good production functions: players continue to differ whilst their valuation of the public good is common.

The coordination problem that arises in a teamwork dilemma leads to three central questions. When will the collective action succeed? If it succeeds, who are the members of the providing team? What role does the size and composition of the contributor pool play? Variants of these questions also arise in the context of more general public-good provision games. Focusing on the role of the contributor pool, this paper offers answers to these questions.

The identification of Nash equilibria of collective-action games cannot provide general answers; neither does the characterization of the Bayesian Nash equilibria of related incomplete-information games. The difficulty faced by static solution concepts is that an equilibrium-selection problem arises whenever the success of a collective action requires teamwork.[2]

Instead, this paper studies a model in which play evolves: a strategy-revision process is considered in which each player periodically chooses a myopic quantal response to the current state of play.[3] The associated long-run distribution over strategy profiles is characterized. When quantal responses approximate myopic best replies the process spends almost all time local to a selected equilibrium of the corresponding complete-information game. While sharp results (Sections 4–5) are available for this case, the analysis may also be applied away from the limit. Using a logit specification applied to a general class of symmetric-valuation games, a complete characterization of long-run behaviour is provided (Section 6).

1. van de Kragt, Orbell and Dawes (1983, p. 113, original emphasis) explained: "Its essentials are very simple: *when the cost of the group's public good is assigned to a subset of members each of whose contribution is necessary to the production of the good, it is reasonable for each to contribute* ... [this logic] does not appear to have been specified in the literature on public-good provision." It may be optimal to rest the fate of a team upon the shoulders of each member; however, this works only if others participate as expected.

2. Palfrey and Rosenthal (1984) characterized mixed-strategy Nash equilibria of a teamwork dilemma in which a subset of the players randomize. Even when pay-offs 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.

3. 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. Diermeier and Van Mieghem (2005*b*) presented a model of consumer boycotts using an approach related to the one taken here.

So why examine a strategy-revision process? One justification is pragmatic: standard solution concepts face an equilibrium-selection problem. One resolution is to investigate the relative robustness of equilibria. Specifying quantal responses is one step towards doing this: noise allows play to move away from Nash equilibria. For instance, in the threshold game there is an equilibrium in which no player contributes; the corresponding quantal-response equilibrium allows for occasional participation against the flow of play. The likelihood of such events can measure the robustness of this, and other, equilibria. Such an approach entails an implicit dynamic story: Section 3 examines this argument in detail. The long-run distribution of play precisely captures the relative robustness of the various states of play.

An empirical justification is that collective-action participation does in fact vary over time. The open-source software movement exemplifies.[4] Recent papers have used data from the SourceForge (sourceforge.net) collaborative development environment to assess the dynamics of voluntary involvement in open-source projects. In one such study, David and Rullani (2006) calculated transition probabilities between different states of participation. Their evidence shows that programmers enter and exit projects regularly. Furthermore, other studies based on SourceForge (Giuri, Ploner, Rullani and Torrisi, 2006, for example) have observed that teamwork is a fundamental organizational feature of the open-source development model.

The dynamics of contribution decisions and the need for teamwork are also present in political settings. For instance, Bulow and Klemperer (1999) documented the passing of President Clinton's U.S. budget in August 1993. The budget required a critical number of supporters in order to pass and hence generate a collective benefit (support for the new Democratic president's first budget). However, the act of support (a participation decision) was privately costly for the legislators owing to the bill's wider unpopularity. Support was in fact built over time as legislators steadily fell into line. The important features highlighted here are the need for teamwork and the dynamic way in which support was built.[5]

The evolution of collective action when success depends upon a critical mass of participation is an important component of sociological phenomena: these features are central to crowd behaviour, the onset of riots and strikes, and other social movements. Granovetter (1978), Oliver and Marwell (1988), Oliver *et al.* (1985), and others, used numerical simulations, experiments, and informal theory to investigate the dynamics of groups involved in collective actions. Such authors have noted that a small number of initial contributions may raise the effectiveness of further contributions and induce mass participation. Heckathorn (1993, 1996), Macy (1990, 1991*a*), and others developed these ideas.

An important determinant of the success of a team-based collective action is the ease with which teams build. However, the relative robustness of established teams is also important. When success of a team depends pivotally on the participation of its members, a collapse can occur when a "bad apple" chooses to depart. Sociologists (Macy, 1991*b*) and psychologists (Colman, 1982, 1995) have noted the importance of bad apples in social dilemmas. Empirical (Barrick, Stewart, Neubert and Mount, 1998; Dunlop and Lee, 2004) and experimental (Haythorn, 1953; Camacho and Paulus, 1995) studies of organizations and teams have found that the presence of bad apples is a strong indicator of poor performance.[6]

The bad-apple phenomenon leads to a further justification for the study of a strategy-revision process. A central message emerging from the formal results presented in this paper is that the

4. Programmers make voluntary contributions to software projects. The "copyleft" licensing regime (Raymond, 1998) ensures that these can then be used freely by all members of the user and programmer community.

5. Bulow and Klemperer (1999) modelled this event as a multi-player war of attrition. In such a game participants (that is, those who concede early on rather than holding out to free ride) anticipate the future behaviour of others. Hence players are more patient than the myopic agents that are the focus of this paper.

6. For an extensive discussion and documentation of such bad-apple effects in the psychology and organizational behaviour literatures, see Felps, Mitchell and Byington (2006).

characteristics of potential contributors can matter even if they do not usually participate in the provision of the public good. Indeed, some of these contributors are bad apples in that their presence can sometimes destablize an otherwise-successful collective action. Such bad apples can enter a successfully operating team, supplant an existing member, and then subsequently leave. This phenomenon relies critically on a dynamic model; an examination of the equilibria of a static game necessarily misses this important feature.

The insights that emerge from the study of such strategy-revision processes are most easily seen in the context of threshold games: a myopic best reply is for a player to contribute if and only if pivotal. When a revising player joins a collective action even when not pivotal the process experiences a low probability "birth". Similarly, when a revising player leaves a successful team then the process experiences a "death". Quantal responses allow the probabilities of births and deaths to respond to pay-offs: a birth is less likely when the cost of participation is high, and a death is less likely when the public good is highly prized. The "noise" in a player's quantal response also matters: under a random-utility interpretation, the birth and death probabilities are higher for more variable pay-offs.

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 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 more likely than a single death.

Who are the members of a successful team? They are (usually) the "most enthusiastic" players who experience the highest birth probabilities. Enthusiasm might stem from low contribution costs, and hence enthusiastic players will be least likely to leave a successful team. Alternatively, enthusiasm may stem from pay-off variability. A player with relatively noisy pay-offs is more likely to join and also more likely to drop out. Such a team member can be a bad apple within the team and cause the collective action to fail. A bad apple within the team can also appear when the players with low contribution costs place little value on the public good. This bad-apple effect sheds light on the role played by the composition of the contributor pool: the success of a collective action is favoured when low contribution costs (high birth probabilities) are associated with high public-good valuations (low death probabilities) so that enthusiasts are hard to kill.

This is not the only instance of the bad-apple effect. A player outside a successfully operating team might destabilize 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.

Moving beyond simple threshold games bad-apple effects remain when a general class of public-good production functions is considered. In this general setting, and even when quantal responses do not approximate best replies, sharp results obtain. When differences in players arise from differences in their costs of participation, the worst kind of bad apple is a player with an intermediate 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, and is optimized when the contributor pool is restricted to a subset of players with sufficiently low contribution costs.

Sections 2 and 3 describe and discuss, respectively, the evolution of collective action, with a particular focus on threshold games. Section 4 presents results for vanishing noise, which are then applied to cases of interest in Section 5. Section 6 expands the scope of the paper by allowing for a completely general class of public-good production technologies, whilst also allowing for non-vanishing noise. The proofs of the propositions in the main text are relegated to Appendix C. These build upon two central theorems: Theorem A1 (Appendix A) completely characterizes the ergodic distribution for vanishing noise; Theorem B1 (Appendix B) relates team success to the association between birth and death probabilities.

2. THE EVOLUTION OF COLLECTIVE ACTION

In this section a collective-action game of interest is described and the mechanism by which play evolves is introduced. A critical discussion of the modelling strategy, an evaluation of other approaches, and an interpretation of the model's components are contained in Section 3.

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 pay-off $u_i(z)$ 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 $Z^* = \{z : z_i = 1 \Leftrightarrow \Delta u_i(z) > 0\}$.

A collective-action game of particular interest is an "n choose m" game 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,

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

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 pay-off 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".[7] The remainder of this section and Sections 4–5 focus exclusively on such dilemmas. However, the insights carry over readily to a fully general class of public-good production functions (Section 6).

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

7. 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 (Myatt and Wallace, 2008).

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 $\Pr[z \to z'] \equiv \Pr[z_{t+1} = z' \,|\, z_t = z]$. The transitions involve steps up and down 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,

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

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

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,

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

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.

Birth and death probabilities differ from each other and across players. As "noise vanishes" (that is, as $\varepsilon \to 0$) these probabilities decline at different rates. Formally,

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

An inspection of (1) reveals that β_i is the (exponential) rate at which b_i vanishes as $\varepsilon \to 0$; it is the "birth cost" of a transition made against the flow of play by a new contributor. Similarly, δ_i is the "death cost" of a step down from Z_m. If birth and death costs differ, then birth and death probabilities vanish at different rates as $\varepsilon \to 0$. This generates state-dependent mutations (Bergin and Lipman, 1996). Sections 4 and 5 require only that (1) holds in the limit as $\varepsilon \to 0$. For the threshold game, Section 6 assumes that (1) holds for all $\varepsilon > 0$. However, Section 6 also considers a wider class of public-good provision games.

3. MODELLING STRATEGY AND OTHER APPROACHES

Firstly, consider the birth and death probabilities introduced in Section 2. One possibility is to set $b_i = d_i = \varepsilon$ for an "error" probability $\varepsilon > 0$. Such a model of mistakes is encompassed by the more general specification employed here. Moreover, state-dependent birth and death probabilities arise naturally from quantal-response strategy revisions (McKelvey and Palfrey, 1995).

If player i is called upon to revise, then a logistic quantal response entails

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

This carries the usual random-utility interpretation, so that birth and death probabilities respond to pay-offs. For instance, if player i's cost of participation is small, then the birth cost β_i is small. To see the logit quantal-response in action, take the Palfrey and Rosenthal (1984) threshold public-good provision game, for which $u_i(z)=\upsilon_i\mathcal{I}[|z|\geq m]-z_i c_i$. Then

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

This specification yields $\beta_i=c_i$ and $\delta_i=\upsilon_i-c_i$, which satisfy (1).

A wider class of quantal-response specifications satisfies (1) in the limit as $\varepsilon\to 0$; this is all that is needed for the characterization of long-run play as $\varepsilon\to 0$ (Sections 4 and 5). Revisions are made by probit quantal-response when player i contributes 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

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

The probability that player i chooses to contribute against the flow of play is now 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 ε, falls to zero. For small ε, 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 $\varepsilon\to 0$,

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

Other quantal-response specifications (see Appendix A) satisfy (5) and so also generate models within the "birth and death cost" framework.[8]

The random-utility interpretation suggests a game of incomplete information and consideration of Bayesian Nash equilibria. Such equilibria do not solve the multiplicity problem: for every Nash equilibrium there is a Bayesian Nash equilibrium of the quantal-response game which purifies the Nash equilibrium. Nevertheless, noisy pay-off realizations ensure that non-Nash profiles are played with non-zero (albeit small) probability. An investigation of these probabilities might yield insight into the relative robustness of the equilibria. However, such an investigation inexorably leads to a dynamic story and hence a strategy-revision process.

To pursue this investigation informally, consider the "team failure" Bayesian Nash equilibrium in which, with high probability, no player contributes.[9] The collective action usually fails; nevertheless, for $\varepsilon>0$ there is a small chance that it succeeds. In fact, for small ε the probability that player i contributes satisfies $\log\Pr[z_i=1]\approx-\beta_i/\varepsilon$. Now, suppose that birth costs differ

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

9. A formal investigation is contained in a not-for-publication appendix (www.restud.com).

and label players so that $\beta_1 < \beta_2 < \cdots < \beta_n$. The probability that the first m players contribute is approximately $b_1 \times b_2 \times \cdots \times b_m$. Since this m-strong subset of players is far more likely to contribute simultaneously than any other subset, for small ε,

$$\log \Pr[|z| \geq m] \approx -\frac{\sum_{i=1}^{m} \beta_i}{\varepsilon}.$$

Heuristically, $\sum_{i=1}^{m} \beta_i$ indexes the robustness of the team-failure equilibrium: it measures the probability that the collective action succeeds despite the fact that it usually fails. Similarly, consider the "team success" Bayesian Nash equilibrium in which, with high probability, the first m players contribute: similar logic reveals that $\min_{i \leq m}[\delta_i]$, which is the death cost of the player most likely to quit, indexes the equilibrium's robustness. If $\min_{i \leq m}[\delta_i] > \sum_{i=1}^{m} \beta_i$, then the team containing the first m players is relatively robust. On the other hand, it is relatively fragile when a "bad apple" team member has a particularly low death cost.

But what of other teams? The bad apple might be replaced with another higher-death-cost player. Consider the m players with the highest death costs. This team of "reliable" players is obtained by excluding bad apples. But would this team ever form? If it consists of those with relatively high birth costs (those reluctant to volunteer), then this seems unlikely.

This informal reasoning suggests a dynamic story. In order to assess the relative robustness of a team it is necessary to think about which team is likely to arise in the first place. A natural candidate is the team containing the first m players, simply because they have the lowest birth costs: this suggests the comparison of $\min_{i \leq m}[\delta_i]$ and $\sum_{i=1}^{m} \beta_i$ made previously. Alas, this comparison neglects the possible role of other feasible teams. This is important because, within a dynamic story, teams can form and collapse via indirect routes.

To see this, begin with a successful team containing the first m players. With some probability (approximately b_j) player $j > m$ volunteers. This does not cause the failure of the collective action; however, it may alter the membership of the team. Once j participates an original team member can drop out. Thus, thinking about how teams form and collapse leads to the study of how the composition of teams may change over time. Now, if player j has a particularly low death cost then the change in membership has reduced the robustness of the team. Player j is a bad apple arising from outside the original team who supplants an existing team member, hastening the subsequent collapse of the collective action.

This informal investigation of the robustness of Bayesian Nash equilibria has led, somewhat inevitably, to a dynamic story. This paper tells the story formally: only in this context can the robustness of the many different equilibria be assessed simultaneously. Furthermore, and crucially, in a dynamic world the bad apples emerging from outside a team are brought to light; they remain hidden when attention is restricted to static Bayesian Nash equilibria.

Returning to the dynamic story considered in this paper, the one-step-at-a-time specification is shorthand for a process in which strategy-revision events arise stochastically. If players are equipped with Poisson alarm clocks, so that individual revision opportunities arrive with a constant hazard rate, then strategy revisions coincide with probability zero. Revising players best reply to the current strategy profile.

This procedure does not necessarily require a great deal of information: revising players need know only whether or not they are pivotal to the collective action. This information is easy to obtain via experimentation: when encountering a revision opportunity a player might test the alternative and switch action only when it generates a higher pay-off.

A second issue that arises is the rationality of the players. A sophisticated decision-maker might anticipate future changes in play by others. This will matter only if the player is sufficiently patient. If players are impatient or if revision opportunities are infrequent, then a fully rational decision-maker is indistinguishable from a myopic one.

A different modelling strategy might allow simultaneous revisions. The (implicit) information requirements are higher: experimentation is more difficult, since (presumably) players would be experimenting simultaneously. More worryingly, a fully rational interpretation is problematic since revising players implicitly assume that others will not change their behaviour and so bounded rationality is essential: allowing revising players to anticipate simultaneous revisions by others would, unfortunately, reintroduce the equilibrium-selection problem.[10]

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

Attention now returns to the one-step-at-a-time strategy-revision process (Section 2). Long-run behaviour is characterized by the ergodic distribution $p_z \equiv \lim_{t\to\infty} \Pr[z_t = z]$. A standard approach (Foster and Young, 1990; Kandori, Mailath and Rob, 1993; Young, 1993) is to examine the properties of this ergodic distribution as $\varepsilon \downarrow 0$; the case of "vanishing noise". When ε 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 1. Allowing noise to vanish ($\varepsilon \downarrow 0$) teamwork succeeds if and only if $\sum_{z\in Z_m} p_z \to 1$, and otherwise it fails. A specific m-strong team $z \in Z_m$ succeeds if and only if $p_z \to 1$.

Characterization of the ergodic distribution is straightforward when the players share the same birth and death costs β and δ. 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 states $|z_t| \in \{0, 1, \cdots, n\}$. Writing $\Pr[Z_k \to Z_{k+1}] \equiv \Pr[z_{t+1} \in Z_{k+1} \,|\, z_t \in Z_k]$, 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 $1-d$. Summarizing,

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

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

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

The transitions characterize a simple birth–death process on $\{0, 1, \cdots, n\}$, 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: $\Pr[Z_k \to Z_{k+1}] \times \sum_{z\in Z_k} p_z = \Pr[Z_{k+1} \to Z_k] \times \sum_{z\in Z_{k+1}} p_z$. Substituting the step-up and step-down transition probabilities from above, the relative likelihood of team success vs. failure is

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

10. An analysis of simultaneous revisions is available in the not-for-publication appendix (www.restud.com).

As $\varepsilon \to 0$, the sign of $\delta - (m-1)\beta$ determines whether this ratio explodes or vanishes.[11] Similar calculations confirm that $\sum_{z \in Z_k} p_z \to 0$ for other layers $k \notin \{0, m\}$. Therefore, a comparison of $(m-1)\beta$ and δ establishes whether the team is successful.

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 (υ 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.[12]

Without loss of generality, players will be labelled in birth-cost order: $\beta_1 < \beta_2 < \cdots < \beta_n$. Thus a player with a lower label i is, heuristically, more enthusiastic. Similarly, for states in Z_m, the team $z^\ddagger \in Z_m$ 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.[13] The literature offers two techniques for the characterization 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.[14]

To use the technique popularized by Kandori *et al.* (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 a 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. The limit set in question is "stochastically stable" (Foster and Young, 1990) 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

11. For $\lim_{\varepsilon \to 0} \left[\sum_{z \in Z_m} p_z / \sum_{z \in Z_0} p_z \right] = \lim_{\varepsilon \to 0} [\exp((\delta - (m-1)\beta)/\varepsilon]$ 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).

12. $\beta_i \neq \beta_j$ and $\delta_i \neq \delta_j$ for all $i \neq j$. These requirements are satisifed 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 $\sum_{z \in Z_0} p_z \to 1$ as $\varepsilon \to 0$.

13. 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 6.

14. 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.

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 A1 (Appendix A) uses tree surgery to characterize 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.[15] Based on these observations, this paper takes a somewhat eclectic approach. Whereas the propositions stated below are derived as corollaries of the aforementioned Theorem A1, the intuition developed throughout the text (formal proofs are relegated to Appendix C) builds upon the radius–coradius methodology.

It proves useful to divide the 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 $\beta_1 < \cdots < \beta_n$) this obtains when death costs are decreasing: $\delta_1 > \cdots > \delta_n$. The members of the most enthusiastic team $z^\ddagger$ are the hardest to kill, and so the candidates for stochastic stability are $z^\ddagger$ 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 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 δ_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 δ_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 β_j) and becomes the

15. For collective-action games there are circumstances in which the tree-surgery approach can provide predictions whereas the radius–coradius analysis cannot; a specific example is provided in footnote 16.

$(m+1)$-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 δ_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 δ_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 minimized 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 minimize this expression generates the easiest indirect route out of $z^\ddagger$, with a total resistance of

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

Comparing direct and indirect routes, the modified coradius of $z^\ddagger$ is simply $\min[\delta_m, \delta^\ddagger]$. Following from this, a sufficient condition for team failure (which can be derived from the radius-coradius theorem) is $\sum_{i=1}^{m-1} \beta_i > \min[\delta_m, \delta^\ddagger]$. 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* $\sum_{i=1}^{m-1} \beta_i < \min[\delta_m, \delta^\ddagger]$*; otherwise teamwork fails.*

Proposition 2 reveals two "bad apple" effects. One possible disruption to the collective action is the departure of a member of $z^\ddagger$; 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 $\delta_m < \delta^\ddagger$, the bad apple (player m) within the team is the weakest link. In contrast, when $\delta_m > \delta^\ddagger$ 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 $\delta_1 < \cdots < \delta_n$. The most enthusiastic team is also the flakiest; the easiest way to disrupt $z^\ddagger$ is for the most enthusiastic player (with the lowest birth cost β_1 and the lowest death cost δ_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 $z^\ddagger$ 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 $z^\ddagger$. However, the death cost of the bad apple within the team rises to δ_2. So, when $\delta_2 - \delta_1 > \beta_m - \beta_1$, is z a better candidate than $z^\ddagger$ 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 summarizes.[16]

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.[17] 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 destabilized, 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 permutation of the same death costs $\hat{\delta}$ "favours team success" if whenever a team succeeded under δ a (possibly different) team certainly will succeed under $\hat{\delta}$.[18] Different permutations of death costs retain the marginal distributions of birth and death costs, but change the joint distribution of their ranks (the empirical copula):

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

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

16. 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 β_{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.

17. 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 A1 (Appendix A) which provides a full characterization of the ergodic distribution for vanishing noise for any death-cost ordering.

18. A permutation of δ is an ordering $\hat{\delta}$ such that for each i, $\hat{\delta}_i = \delta_j$ for some j, (and vice versa).

Concordance provides a (partial) ordering over copulae. Equivalently, since birth costs are arranged in size order, it is a partial ordering over death-cost permutations. A more concordant permutation shifts low death costs towards the players with low birth costs.[19]

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 5 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 6; bad-apple effects remain critical.

5. ON THE SUCCESS OF A COLLECTIVE ACTION

Propositions 1–4 take as primitive the somewhat abstract pattern of birth and death costs. This section returns to the more focused setting of a threshold public-good provision game.

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 $u_i(z) = \upsilon \times \mathcal{I}[|z| \geq m] - z_i c_i$, where players are ordered so that $0 < c_1 < \cdots < c_n < \upsilon$. As noted in Section 2, this specification generates a contribution incentive

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

Given that $c_i < \upsilon$, 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 pay-offs, 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 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 $\upsilon > \sum_{i=1}^{m} c_i$.

Definition 3. A symmetric-valuation collective action is privately feasible $\Leftrightarrow \upsilon > \sum_{i=1}^{m} c_i$.

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 internalization of positive externalities works against public-good provision.

19. The set of permutations form a lattice with maximal and minimal members which correspond to perfect concordance and perfect discordance, respectively. An increase in concordance implies an increase in the standard empirical measures of association, such as Spearman's ρ and Kendall's τ.

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 $\delta_i = \upsilon - c_i$. 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 $\sum_{i=1}^{m-1} \beta_i < \min[\delta_m, \delta^\ddagger]$. When $\delta_m < \delta^\ddagger$, 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 $\sum_{i=1}^{m-1} c_i < \upsilon - c_m$, or equivalently $\sum_{i=1}^{m} c_i < \upsilon$. This, of course, is the criterion for private feasibility.

In 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 $\beta_j + \delta_j = \upsilon$, and so

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

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. Summarizing:

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.*

Proposition 5 reveals characteristics of the contributor pool 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 lowest-cost contributors, and variation in provision costs helps to make the best even better.

Proposition 5 restricts to a scenario in which players share the same valuation for the public good. Variation in υ_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. 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 minimizes disruption and so favours success. Alternatively, the public good may be of little use to enthusiasts. This may lead to concordant birth and death costs, and hence (Proposition 4) this favours failure.[20]

Proposition 5 reveals that a riskier distribution of contribution costs helps private feasibility and team success. But what of the distribution of valuations? Suppose that contribution costs are negatively related to valuations, so that enthusiastic players prize the public good. The bad apple in a successful team of enthusiasts is player m, with the m-th highest valuation and a death cost of $\delta_m = \upsilon_m - c_m$. If $m < n/2$ then this player's valuation is above average, in the sense that it

20. For open-source software (an example noted in Section 1) the first scenario might correspond to a networking utility, which is useful for skilled programmers who are best able to contribute to its development. The second scenario might apply to a word processor; skilled programmers (the enthusiasts) might prefer a simple text editor, and so place little value on the public good. These examples of positive and negative correlation were also identified by Johnson (2002) who modelled software provision as a volunteer's dilemma.

is above the median. 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 δ_m; this favours team success. 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 realized 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 pay-offs, 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 ξ_i^2. Hence players with the lowest birth costs are enthusiastic not because they are efficient but rather because their pay-offs are particularly idiosyncratic; a birth-cost ordering leads to $\xi_1^2 > \cdots > \xi_n^2$. 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.[21] Furthermore, the presence of a player with a particularly high pay-off variance can result in team failure: collective 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 δ of a player is determined by the variance of $\tilde{v}_i - \tilde{c}_i$. This is, in turn, affected by any "within" correlation between valuation and contribution cost. For instance, if

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

then ρ is the correlation coefficient between the random-utility valuation and contribution-cost realizations of a player. An increase in this correlation increases the death cost δ, 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 realizations of each individual reduces death costs and so favours team success.*

21. 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

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

The parameter λ_i then indexes the idiosyncrasy of player i's noisy pay-offs.

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 Propositions 4 and 7. Looking across the player set (the "between" dimension of a panel) negative correlation creates the desirable discordance between birth and death costs; enthusiasts are reliable, and this favours team success. In contrast, looking at the pay-off realizations of a 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.

6. WELFARE AND THE CONTRIBUTOR POOL

This section studies evolving play when noise is bounded away from zero, so that quantal-response strategy revisions do not necessarily approximate best-replies. This is desirable because the stochastic-stability results presented in earlier sections omit important factors. For instance, and as noted in Section 4, 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 ε is fixed and n is allowed to rise, then an inspection of (6) reveals that the relative likelihood of team success vs. failure can grow without bound: there is a larger pool of potential contributors who can help to build a successful team.

Progress is facilitated by restricting to play that evolves via logit quantal response, so that

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

and by assuming that all players value the public good symmetrically. For the threshold game considered previously, this means that $v_i = v$ for all i. Fortunately, the techniques employed here permit consideration of a more general class of public-good provision games in which production is a general increasing function $G(\cdot)$ of the number of contributions:

$$u_i(z) = G(|z|) - c_i z_i.$$

The threshold game studied so far is obtained by setting $G(|z|) = v \times \mathcal{I}[|z| \geq m]$. With symmetric valuations, birth and death costs in a threshold game satisfy $\beta_i = c_i$ 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 if the collective action is privately feasible. The notion of private feasibility is also useful when quantal-response noise is retained and when a general form for $G(\cdot)$ is permitted.

Definition 4. For a symmetric-valuation public-good provision game evolving via logit quantal response, the private feasibility of a state z is $\psi(z) = G(|z|) - \sum_{i=1}^{n} c_i z_i$.

This index of private feasibility comprises a private valuation $G(|z|)$ for the public good 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,

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

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 $p_z \times \Pr[z \to z'] = p_{z'} \times \Pr[z' \to z]$ for all such pairs.[22] 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,

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

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.[23]

Proposition 8. *When the play of a symmetric-valuation public-good provision game with a general production function $G(\cdot)$ evolves via logit quantal response,*

$$p_z = \frac{\exp[\psi(z)/\varepsilon]}{\sum_{z' \in Z} \exp[\psi(z')/\varepsilon]} \quad \textit{where} \quad \psi(z) = G(|z|) - \sum_{i=1}^{n} c_i z_i.$$

Hence the long-run probability of a state is determined by its private feasibility.

Setting $G(0) = 0$ the no-contribution state in Z_0 has a potential of zero. If $G(|z|) < \sum_{i=1}^{n} c_i z_i$ for all $|z| > 0$ then, heuristically, the collective action is privately infeasible. 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 internalize the externalities of provision.

Attention now returns to the bad-apple effects that were central to Propositions 1–7. In the context of a threshold game, a bad apple is either (i) the weak link within the most enthusiastic team; or (ii) a player from outside the most enthusiastic team who destabilizes the team by supplanting a more reliable team member. For the more general provision games considered here, a bad apple is 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 successful collective action 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 pay-offs: $w(z) \equiv \sum_{i=1}^{n} u_i(z) = nG(|z|) - \sum_{i=1}^{n} z_i c_i$. In the long run, z is played with probability p_z, and hence (long run) expected welfare is

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

22. 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.

23. Proposition 8 applies to a more general class of public-good provision games (Myatt and Wallace, 2007). Allow each player's action z_i to be chosen from a finite set at a private cost of $c_i(z_i)$, and specify production $G(z)$ to be a general function of the entire strategy profile. Set a player's pay-off 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.

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 $(n-1)G(|z|)$. Proposition 9 confirms that bad apples hurt welfare.[24]

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. Summarizing: 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.[25]

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 supplant more efficient contributors). For the latter case, allowing $c_i \to \infty$ is equivalent to excluding player i from the contributor pool. A straightforward corollary of Proposition 9 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 i from volunteering. This argument suggests a rationale for limiting the size of the contributor pool; this will serve to exclude the unwanted bad apples. A natural question suggests itself: 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-maximizing contributor pool consists of the most enthusiastic n^* players. Specializing to a threshold game, either $n^* = 0$ or $n^* \geq m$. If the collective action is privately feasible, then $n^* = m$ for ε sufficiently small.*

The final claims concern threshold games: when the collective action is privately feasible $z^\ddagger$ is stochastically stable, and so $W \to w(z^\ddagger)$ as noise vanishes so long as the members of $z^\ddagger$ can participate. Now, $w(z^\ddagger) - W_i^+ \geq c_{m+1} - c_m > 0$ for all $i > m$, and hence (for small ε) 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

24. Proposition 9 holds when production is a general function $G(z)$ of the contribution profile.

25. The same logic was used by Myatt and Wallace (2007, Proposition 1) in a paper which provides an evolutionary justification for the use of thresholds in collective-action problems.

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. For a threshold game, this comparative-static exercise corresponds 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 a critical mass required for success (for threshold games, it may be that $n^* > m$ for larger ε) 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 $\varepsilon \to 0$. 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 $\mathcal{E} = -\lim_{\varepsilon \downarrow 0} \varepsilon \log g(\varepsilon)$ 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 $\varepsilon \to 0$. For a set $\{p_l(\varepsilon)\}$ with exponential costs $\{\mathcal{E}_l\}$,

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

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 $\delta_i = \upsilon_i - c_i$. 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*

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

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* $\delta_i = (\upsilon_i - c_i)^2/(2\gamma_i^2)$.

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

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

where $\phi(\cdot)$ is the density of the standard normal. The penultimate equality follows from an application of l'Hôpital's rule as $x \to \infty$, 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. ‖

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 8, suppose that the noise in $\Delta\tilde{u}_i(z)$ is drawn from the generalized error distribution (equivalently, the exponential power distribution). This has a density $f(x) \propto \exp(-|x|^\nu)$, where ν 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 $\mathcal{E}_{zz'} \equiv \mathcal{E}(\Pr[z \to z'])$, 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,*

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

For $\varepsilon > 0$, there is a unique ergodic distribution $p = \{p_z\}_{z\in Z}$. A graph-theoretic technique will be used to characterize p as $\varepsilon \to 0$. 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 $q_z = \sum_{h\in H_z} \prod_{(s,s')\in h} \Pr[s \to s']$.*

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 $\varepsilon \to 0$ 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),

$$\mathcal{E}(q_z) = \mathcal{E}\left(\sum\nolimits_{h\in H_z} \prod\nolimits_{(s,s')\in h} \Pr[s \to s']\right) = \min_{h\in H_z} \mathcal{E}\left(\prod\nolimits_{(s,s')\in h} \Pr[s \to s']\right) = \min_{h\in H_z} \mathcal{E}_h .$$

From (8), $\mathcal{E}(q_z) < \mathcal{E}(q_{z'}) \Rightarrow \lim_{\varepsilon\to 0}[q_{z'}/q_z] = 0$, 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 $\varepsilon \to 0$; they are stochastically stable.

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

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

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 $\beta_1 < \cdots < \beta_n$. The "most enthusiastic team" is $z^\ddagger \equiv \{z \in Z_m : z_i = 1 \Leftrightarrow i \le m\}$. Define $\kappa \equiv \arg\min_{i<m}[\delta_i]$ and $\iota \equiv \arg\min_{i\le m}[\delta_i]$; 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 $Z^\diamond = \{z \in Z_m : z_i = 1\, \forall\, i < m$ 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 A1. *If $\delta_\iota \ge \min[\delta_\mu, \beta_\iota]$ and*

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

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

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

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

so teamwork succeeds but without player m. Otherwise $\mathcal{E}(z^0) \le \mathcal{E}(z)$ 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

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

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 $z^\ddagger$ satisfies*

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

Proof. Consider a tree rooted at $z^{\ddagger}$. There is no exit from $z^{\ddagger}$, but exits are required for all $z \in Z_m$ where $z \neq z^{\ddagger}$. Total costs of exits from Z_m are therefore at least $A - \min_{i\in N}[z_i^{\ddagger}\delta_i + (1-z_i^{\ddagger})\beta_i]$. 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 $z^{\ddagger} \in Z_m$. Hence

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

This expression provides a lower bound to the exponential cost of any tree rooted at $z^{\ddagger}$. 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 $z^{\ddagger}$ comes at a cost of β_{m+1}, since $z_i^{\ddagger} = 1$ for all $i \leq m$, and $z^{\ddagger} = 0$ for all $i > m$. Similarly, the cheapest death out of $z^{\ddagger}$ comes at a cost of δ_ι. Hence $\min_{i\in N}[z_i^{\ddagger}\delta_i + (1-z_i^{\ddagger})\beta_i]$ in (10) can be replaced by $\min[\delta_\iota, \beta_{m+1}]$, yielding $\mathcal{E}(z^{\ddagger}) \geq A - \min[\delta_\iota, \beta_{m+1}] + \sum_{i=1}^{m-1}\beta_i$.

It is now shown that transitions from remaining states may be constructed at no additional exponential cost, so that a tree rooted at $z^{\ddagger}$ is obtained that attains the lower bound in (10). To do this, take a state $z \in Z_m$ where $z \neq z^{\ddagger}$ such that $\min_{i\in N}\left[z_i\delta_i + (1-z_i)\beta_i\right] = \delta_j$ 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 $z^{\ddagger}$. In this case, there is a path from z into $z^{\ddagger}$. A second possibility is that a zero-cost step up might be taken directly into $z^{\ddagger}$. 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 $z^{\ddagger}$. 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 $z^{\ddagger}$.

Next, take a state $z \in Z_m$ where $z \neq z^{\ddagger}$ 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 $z \neq z^{\ddagger}$, so $\beta_k > \beta_j$.[26] 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 $z^{\ddagger} = \arg\min_{z\in Z_m} B(z)$ must be reached. Consequently there are no cycles, and there is a path from every state in Z_m to $z^{\ddagger}$. 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. ‖

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

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

Proof. Follow the argument leading to (10) in the proof of Lemma 5. ‖

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*

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

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 $z \neq z^{\ddagger}$ follow the procedure described in the proof of Lemma 5 to construct a sequence of zero-cost transitions that connect each such state to $z^{\ddagger}$. 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 $z^{\ddagger}$. 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.

26. This is where the genericity assumption (see footnote 12) 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.

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 $z^{\ddagger}$. 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 $z^{\diamond}_j = 1$ must die at a cost of δ_j to move into layer Z_{m-1}. (In preparation for step (v) below, note that it might well be that $z^{\diamond} = z^{\ddagger}$.)

Step (iv). Take the set of states $z \in Z_m$ that are *not* encountered on the path from $z^{\ddagger}$ 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 $z^{\ddagger}$ to $z^{\diamond}$.

Step (v). Suppose that $z^{\ddagger} = z^{\diamond}$. 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 δ_ι among the team members at $z^{\ddagger}$. Relative to the benchmark, there is no exit upward. The exponential cost of this tree is

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

Step (vi). Suppose instead that $z^{\ddagger} \neq z^{\diamond}$. Write $\tilde{N} = \{i : z^{\diamond}_i = 1 \text{ and } z^{\ddagger}_i = 0\}$ for the contributors acquired along the path from $z^{\ddagger}$ 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 were 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 $z^{\ddagger}$ 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 β_j and δ_j and deductions of β_{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 $z^{\ddagger}$. 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 $z^{\diamond} \neq z^{\ddagger}$ 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

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

which (given the definition of $\delta^{\ddagger}$) is the required expression in the statement of the lemma. ||

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 $A - \min_{i\in N}[z^j_i \delta_i + (1 - z^j_i)\beta_i] + \sum_{i=1}^{m-1} \beta_i$. The team z^j contains j, therefore an element of the second term is δ_j. It does not contain m, but does contain each $i < m$. Therefore the lowest birth cost such that $z^j_i = 0$ is β_m. Finally, the lowest death cost may be δ_j or it may be the lowest death cost of the first $m-1$ players, δ_κ. 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 $z^{\ddagger}$ 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 $z^{\ddagger}$ leading to z^j. The cheapest such branch has cost $\min[\beta_{m+1}, \delta_\iota]$. But

$\delta_\imath < \min[\max_{i>m}\delta_i, \beta_\imath]$ and $\imath = m$. Hence $\delta_\imath < \beta_\imath = \beta_m < \beta_{m+1}$. The cheapest exit from $z^\ddagger$ is to kill player m (at the cost of $\delta_\imath$). Adding player j at zero cost from the resultant state in Z_{m-1} leads directly to z^j. ‖

Proof of Theorem A1 *for* $\delta_\imath \geq \min[\delta_\mu, \beta_\imath]$. A first claim is that $\min\{\mathcal{E}(z^0), \mathcal{E}(z^\ddagger)\} \leq \mathcal{E}(z)$ 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 $\mathcal{E}(z^\ddagger) \leq \mathcal{E}(z)$ for $z \in Z_m$. Using Lemmas 5 and 6, it is sufficient to show

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

since the L.H.S. is $\mathcal{E}(z^\ddagger)$ and the R.H.S. is a lower bound on $\mathcal{E}(z)$ for $z \in Z_m$. Implying (12) it will be shown that $\min_{i\in N}[z_i\delta_i + (1-z_i)\beta_i] \leq \min[\delta_\imath, \beta_{m+1}]$. Three subcases are considered.

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

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

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

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

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

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

(ii) Now consider the alternative $\delta_\imath > \beta_{m+1}$. From Lemma 5, $\mathcal{E}(z^\ddagger) = A - \beta_{m+1} + \sum_{i=1}^{m-1}\beta_i$, and direct application of Lemma 7 verifies (13). This proves the first statement in Theorem A1.

Proof of Theorem A1 *for* $\delta_\imath < \min[\delta_\mu, \beta_\imath]$. 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, $\delta_\imath < \beta_\imath < \beta_{m+1}$. Thus, using Lemma 7, $\mathcal{E}(z^0) = A$. Using Lemma 6, a sufficient condition for z^0 to be selected is

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

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

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

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

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

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

The remaining case is when $\delta_\imath < \min[\delta_\mu, \beta_\imath]$ and $\imath = m$. 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,

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

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 $z^{\ddagger}$ and the members of $Z^{\diamond}$. For $z^j \in Z^{\diamond}$, Lemma 8 applies. Now, $z^{\ddagger}$ is a candidate for selection only if $\mathcal{E}(z^{\ddagger}) \leq \mathcal{E}(z^j)$ for all $z^j \in Z^{\diamond}$. 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 $z^{\ddagger}$ is not selected.

This leaves only z^0 and teams $z^j \in Z^{\diamond}$ 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 $\sum_{i=1}^{m-1} \beta_i < \min\left[\delta_\mu, \delta_\kappa, \beta_m\right]$ 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 δ_κ and β_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 $z \in Z^{\diamond}$ where $z \neq z^\mu$. On the other hand, if $\delta_\mu > \min[\delta_\kappa, \beta_m]$, then there are (potentially) many teams in $Z^{\diamond}$ 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^μ. ‖

A careful reading of the proof of Theorem A1 reveals that for generic (Footnote 12) 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 $\sum_{z \in Z_0} p_z \to 1$ as $\varepsilon \to 0$. 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 $\varepsilon \to 0$. 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 permutation 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 permutation of δ 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 δ_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 δ, there are k death costs within the first m players lower than δ_μ. Every player $j > m$ has a death cost lower than δ_μ. Therefore, given that (at least) one of the death costs *above* δ_μ 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 $\hat{C}$. ‖

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 B1. *If $\hat{\delta}$ is a discordant permutation of δ, then $\mathcal{E}(z) < \mathcal{E}(z^0) \Rightarrow \hat{\mathcal{E}}(z') < \hat{\mathcal{E}}(z^0)$ for $z, z' \in Z$.*

Proof of Theorem B1. Clearly, either $z = z^{\ddagger}$ or $z \in Z^\mu$. Suppose $z = z^{\ddagger}$ and $\delta_\iota \leq \beta_{m+1}$. Then $\mathcal{E}(z^0) = A$ and $\mathcal{E}(z^{\ddagger}) = A + \sum_{i=1}^{m-1} \beta_i - \delta_\iota$ from Lemmas 7 and 5. Now $\mathcal{E}(z^{\ddagger}) < \mathcal{E}(z^0) \Rightarrow \sum_{i=1}^{m-1} \beta_i < \delta_\iota$. By Lemma 9, $\hat{\delta}_{\hat{\iota}} \geq \delta_\iota > \sum_{i=1}^{m-1} \beta_i$. If $\hat{\delta}_{\hat{\iota}} \leq \beta_{m+1}$ then $\hat{\mathcal{E}}(z^{\ddagger}) - \hat{\mathcal{E}}(z^0) = \sum_{i=1}^{m-1} \beta_i - \hat{\delta}_{\hat{\iota}} < 0$ and the result follows. If $\hat{\delta}_{\hat{\iota}} > \beta_{m+1}$ then (again from Lemmas 7 and 5),

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

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

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

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

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

$$\min[\delta^{\ddagger}, \delta_\imath] \le \delta_\imath \equiv \min_{j>m} \left[\beta_j + \delta_j - \min(\beta_m, \delta_j)\right] \le \beta_k + \hat{\delta}_k - \min[\beta_m, \hat{\delta}_k] = \min[\hat{\delta}^{\ddagger}, \hat{\delta}_{\hat{\imath}}], \tag{15}$$

as required. Suppose $\delta_k > \hat{\delta}_k$. Now $\hat{C}(k, \hat{r}(k)) = \sum_{i=1}^{k} \mathcal{I}[\hat{\delta}_i \le \hat{\delta}_k] = 1$. To see this note that were there a $j > m$ such that $\hat{\delta}_j \le \hat{\delta}_k$ then k could not have been the minimizer in the first place. If there were a $j \le m$ such that $\hat{\delta}_j \le \hat{\delta}_k$ then certainly $\hat{\delta}_{\hat{\imath}} \le \hat{\delta}_k$, and thus $\hat{\delta}_{\hat{\imath}}$ 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 permutation of δ, $C(k, \hat{r}(k)) = \sum_{i=1}^{k} \mathcal{I}[\delta_i \le \hat{\delta}_k] \ge 1$. So there exists $i \le k$ such that $\delta_i \le \hat{\delta}_k$, $i \ne k$, since $\delta_k > \hat{\delta}_k$ by assumption. Thus there exists $i < k$ where $\delta_i \le \hat{\delta}_k$. For such an $i > m$: $\beta_i + \delta_i - \min[\beta_m, \delta_i] \le \beta_k + \hat{\delta}_k - \min[\beta_m, \hat{\delta}_k]$; (15) holds with the second inequality strict. For such an $i \le m$: $\delta_\imath \le \delta_i \le \hat{\delta}_k < \beta_k + \hat{\delta}_k - \min[\beta_m, \hat{\delta}_k]$, and hence $\min[\delta^{\ddagger}, \delta_\imath] < \beta_k + \hat{\delta}_k - \min[\beta_m, \hat{\delta}_k]$ as required.

Suppose now that $z \ne z^{\ddagger}$ but $z \in Z^{\mu}$. Then, by Theorem A1, $\delta_\imath < \min[\delta_\mu, \beta_\imath]$, $\imath = m$, and $\sum_{i=1}^{m-1} \beta_i < \min[\delta_\kappa, \delta_\mu, \beta_m]$. First, suppose that $\hat{\imath} \ne m$. Consider $z' = z^{\ddagger}$. If $\hat{\delta}_{\hat{\imath}} \le \beta_{m+1}$ then $\hat{\mathcal{E}}(z^{\ddagger}) - \hat{\mathcal{E}}(z^0) = \sum_{i=1}^{m-1} \beta_i - \hat{\delta}_{\hat{\imath}}$ by Lemmas 5 and 7. Since $\hat{\imath} < m$: $\hat{\delta}_{\hat{\imath}} = \hat{\delta}_{\hat{\kappa}} \ge \delta_\kappa \ge \min[\delta_\kappa, \delta_\mu, \beta_m] > \sum_{i=1}^{m-1} \beta_i$ (the first inequality follows from Lemma 9) and so $\hat{\mathcal{E}}(z') = \hat{\mathcal{E}}(z^{\ddagger}) < \hat{\mathcal{E}}(z^0)$ as required. If $\hat{\delta}_{\hat{\imath}} > \beta_{m+1}$, then $\hat{\mathcal{E}}(z^{\ddagger}) - \hat{\mathcal{E}}(z^0) = \sum_{i=1}^{m-1} \beta_i - \min[\hat{\delta}^{\ddagger}, \hat{\delta}_{\hat{\imath}}]$ again by Lemmas 5 and 7. The minimum in this expression is at least β_{m+1}, and $\beta_{m+1} > \beta_m \ge \min[\delta_\kappa, \delta_\mu, \beta_m] > \sum_{i=1}^{m-1} \beta_i$. Once again $\hat{\mathcal{E}}(z') = \hat{\mathcal{E}}(z^{\ddagger}) < \hat{\mathcal{E}}(z^0)$ as required.[27] The remaining case is when $\hat{\imath} = m$.

First, suppose that $\hat{\delta}_{\hat{\imath}} < \min[\beta_{\hat{\imath}}, \hat{\delta}_{\hat{\mu}}]$. 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 A1). Suppose this inequality breaks. Noting that $\hat{\delta}_{\hat{\kappa}} \ge \delta_\kappa$, this can only happen if $\hat{\delta}_{\hat{\imath}} < \hat{\delta}_{\hat{\mu}} \le \sum_{i=1}^{m-1} \beta_i < \min[\beta_m, \hat{\delta}_{\hat{\kappa}}]$. So $\hat{\delta}_{\hat{\mu}} < \delta_\mu$. Thus $\delta_\mu = \hat{\delta}_i$ for some $i \le m$. Then there must exist some $j > m$ such that $\hat{\delta}_j = \delta_k$ for $k \le m$. If $k < m$, then $\hat{\delta}_{\hat{\mu}} \ge \delta_k \ge \delta_\kappa > \sum_{i=1}^{m-1} \beta_i$: a contradiction. If $k = m$, then suppose that $i = m$ so that $\hat{\delta}_{\hat{\imath}} = \delta_\mu > \hat{\delta}_{\hat{\mu}}$, a contradiction. Alternatively, suppose that $\hat{\delta}_i = \delta_\mu$, some $i < m$. Then $\delta_i \ge \delta_\kappa$. If $\hat{\delta}_j = \delta_i$ with $j > m$, then once again $\hat{\delta}_{\hat{\mu}} \ge \delta_i \ge \delta_\kappa > \sum_{i=1}^{m-1} \beta_i$. If $\hat{\delta}_j = \delta_i$ with $j = m$, then $\hat{\delta}_{\hat{\imath}} = \delta_i \ge \delta_\kappa$, but $\hat{\delta}_{\hat{\mu}} > \hat{\delta}_{\hat{\imath}} \ge \delta_\kappa > \sum_{i=1}^{m-1} \beta_i$: a final contradiction.

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

APPENDIX C. OMITTED PROOFS

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

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

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

27. Since $\imath = m$ is a requirement for $z \in Z^{\mu}$ to be selected, it must therefore follow that $z^{\ddagger}$ is selected if $\hat{\imath} \ne m$. This in turn means that the precondition $\hat{\delta}_{\hat{\imath}} \ge \min[\beta_{\hat{\imath}}, \hat{\delta}_{\hat{\mu}}]$ is met (see Theorem A1). Of course, this is so, since $\delta_\kappa > \sum_{i=1}^{m-1} \beta_i \Rightarrow \delta_\kappa > \beta_\kappa = \beta_{\hat{\imath}} \ge \min[\beta_{\hat{\imath}}, \hat{\delta}_{\hat{\mu}}]$ and $\hat{\delta}_{\hat{\imath}} = \hat{\delta}_{\hat{\kappa}} \ge \delta_\kappa$.

Proof of Proposition 2. When $\delta_1 > \cdots > \delta_n$, $\iota = m$ and $\mu = m+1$. The first part of Theorem A1 applies, since $\delta_m \geq \min[\delta_{m+1}, \beta_m]$ automatically. Thus $\sum_{i=1}^{m-1} \beta_i < \min[\delta^{\ddagger}, \delta_m]$ implies $\mathcal{E}(z^{\ddagger}) < \mathcal{E}(z)$ for all $z \neq z^{\ddagger}$ and otherwise $\mathcal{E}(z^0) < \mathcal{E}(z)$ for all $z \neq z^0$, as required. ‖

Proof of Proposition 3. When $\delta_1 < \cdots < \delta_n$, $\iota = 1$ 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 A1. 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 $\min[\delta^{\ddagger}, \delta_\iota] = \delta_1$. To see this latter fact, note that $\delta^{\ddagger} \equiv \min_{j>m}[\beta_j + \delta_j - \min[\delta_j, \beta_m]] \geq \delta_{m+1} > \delta_1$. Thus, applying Theorem A1 again gives

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

for all $z \neq z^{\ddagger}$. 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 $z^{\ddagger}$; its reverse sufficient for selection of z^0. ‖

Proof of Proposition 4. From the application of Theorems A1 and B1. ‖

Proof of Propositions 5–8. From arguments given in the main text. ‖

Proof of Proposition 9. $w(z) = nG(|z|) - \sum_{i=1}^{n} z_i c_i$ and $\psi(z) = G(|z|) - \sum_{i=1}^{n} z_i c_i$. 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

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

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

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

At a stationary point the second derivative is strictly positive and so W is strictly quasi-convex in c_i. The R.H.S. of (17) is strictly positive for c_i sufficiently large and strictly negative for c_i sufficiently small (and possibly negative). ‖

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

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

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 first 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 (owing 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 suboptimal; a contradiction. This means that the optimal contributor pool may be empty, or it may be $\{1, \cdots, n^*\}$ for some n^*. For a threshold game, if $n^* < m$, 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 follows from the argument given in the text. ‖

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.

REFERENCES

BARRICK, M. R., STEWART, G. L., NEUBERT, M. J. and MOUNT, M. K. (1998), "Relating Member Ability and Personality to Work-Team Processes and Team Effectiveness", *Journal of Applied Psychology*, **83** (3), 377–391.

BENDOR, J., DIERMEIER, D. and TING, M. (2003), "A Behavioral Model of Turnout", *American Political Science Review*, **97** (2), 261–280.

BERGIN, J. and LIPMAN, B. (1996), "Evolution with State-Dependent Mutations", *Econometrica*, **64** (4), 943– 956.

BERGSTROM, T. C., BLUME, L. and VARIAN, H. R. (1986), "On the Private Provision of Public Goods", *Journal of Public Economics*, **29** (1), 25–49.

BERK, R. (1974), "A Gaming Approach to Crowd Behavior", *American Sociological Review*, **39** (3), 355–373.

BLUME, L. E. (1995), "The Statistical Mechanics of Best-Response Strategy Revision", *Games and Economic Behavior*, **11** (2), 111–145.

BLUME, L. E. (1997), "Population Games", in W. B. Arthur, S. N. Durlauf and D. A. Lane (eds.) *The Economy as an Evolving Complex System II* (Reading, MA: Addison-Wesley) 425–460.

BLUME, L. E. (2003), "How Noise Matters", *Games and Economic Behavior*, **44** (2), 251–271.

BLUME, L. E. and DURLAUF, S. N. (2001), "The Interactions-Based Approach to Socioeconomic Behaviour", in S. N. Durlauf and H. P. Young (eds.) *Social Dynamics* (Cambridge, MA: MIT Press) 15–44.

BULOW, J. and KLEMPERER, P. D. (1999), "The Generalized War of Attrition", *American Economic Review*, **89** (1), 175–189.

CAMACHO, L. M. and PAULUS, P. B. (1995), "The Role of Social Anxiousness in Group Brainstorming", *Journal of Personality and Social Psychology*, **68** (6), 1071–1080.

COLMAN, A. M. (1982), *Game Theory and Experimental Games: The Study of Strategic Interaction* (Oxford: Pergamon).

COLMAN, A. M. (1995), *Game Theory and its Applications in the Social and Biological Sciences*, 2nd edn (Oxford: Butterworth-Heinemann).

CORNES, R. and SANDLER, T. (1984), "Easy Riders, Joint Production, and Public Goods", *The Economic Journal*, **94** (375), 580–598.

DAVID, P. A. and RULLANI, F. (2006), "Micro-dynamics of Free and Open-Source Software Development: Lurking, Laboring, and Launching New Projects on SourceForge" (LEM Working Paper No. 2006/26, Sant' Anna School of Advanced Studies).

DIERMEIER, D. and VAN MIEGHEM, J. A. (2005*a*), "Coordination and Turnout in Large Elections" (Mimeo, Northwestern University).

DIERMEIER, D. and VAN MIEGHEM, J. A. (2005*b*), "A Stochastic Model of Consumer Boycotts" (Mimeo, Northwestern University).

DUNLOP, P. D. and LEE K. (2004), "Workplace Deviance, Organizational Citizenship Behavior, and Business Unit Performance: the Bad Apples Do Spoil the Barrel", *Journal of Organizational Behavior*, **25** (1), 67–80.

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.

FELPS, W., MITCHELL,T. R. and BYINGTON, E. (2006), "How, When, and Why Bad Apples Spoil the Barrel: Negative Group Members and Dysfunctional Groups", in B. Staw (ed.) *Research in Organizational Behavior*, Vol. 27 (Amsterdam: Elsevier) 181–230.

FOSTER, D. and YOUNG, H. P. (1990), "Stochastic Evolutionary Game Dynamics", *Theoretical Population Biology*, **38** (2), 219–232.

FREIDLIN, M. I. and WENTZELL, A. D. (1998), *Random Perturbations of Dynamical Systems*, 2nd edn (Berlin/New York: Springer-Verlag).

GIURI, P., PLONER, M., RULLANI, F. and TORRISI, S. (2006), "Skills, Division of Labor, and Performance in Collective Inventions: Evidence from Open-Source Software" (LEM Working Paper No. 2004/19, Sant'Anna School of Advanced Studies).

GRANOVETTER, M. (1978), "Threshold Models of Collective Behavior", *American Journal of Sociology*, **83** (6), 1420–1443.

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* (Oxford: Philip Allan).

HAYTHORN, W. (1953), "The Influence of Individual Members on the Characteristics of Small Groups", *Journal of Abnormal and Social Psychology*, **48** (2), 276–284.

HECKATHORN, D. D. (1993), "Collective Action and Group Heterogeneity: Voluntary Provision Versus Selective Incentives", *American Sociological Review*, **58** (3), 329–350.

HECKATHORN, D. D. (1996), "The Dynamics and Dilemmas of Collective Action", *American Sociological Review*, **61** (2), 250–277.

INNES, R. (2006), "A Theory of Consumer Boycotts under Symmetric Information and Imperfect Competition", *Economic Journal*, **116** (511), 355–381.

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., MAILATH, G. J. and ROB, R. (1993), "Learning, Mutation and Long-Run Equilibria in Games", *Econometrica*, **61** (1), 29–56.

MACY, M. W. (1990), "Learning Theory and the Logic of Critical Mass", *American Sociological Review*, **55** (6), 809–826.

MACY, M. W. (1991*a*), "Chains of Cooperation: Threshold Effects in Collective Action", *American Sociological Review*, **56** (6),730–747.

MACY, M. W. (1991*b*), "Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange", *American Journal of Sociology*, **97** (3), 808–843.

MARWELL, G., OLIVER, P. E. and PRAHL, R. (1988), "Social Networks and Collective Action: A Theory of the Critical Mass. III", *American Journal of Sociology*, **94** (3), 502–534.

MCKELVEY, R. D. and PALFREY, T. R. (1995), "Quantal Response Equilibria for Normal Form Games", *Games and Economic Behavior*, **10** (1), 6–38.

MONDERER, D. and SHAPLEY, L. S. (1996), "Potential Games", *Games and Economic Behavior*, **14** (1), 124–143.

MYATT, D. P. and WALLACE, C. (2003), "A Multinomial Probit Model of Stochastic Evolution", *Journal of Economic Theory*, **113** (2), 286–301.

MYATT, D. P. and WALLACE, C. (2007), "Evolution Teamwork and Collective Action: Production Targets in the Private Provision of Public Goods" (Mimeo, University of Oxford).

MYATT, D. P. and WALLACE, C. (2008), "An Evolutionary Analysis of the Volunteer's Dilemma", *Games and Economic Behavior* (in press).

OLIVER, P. E. and MARWELL, G. (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. and MARWELL, G. (2001), "Whatever Happened to Critical Mass Theory? A Retrospective and Assessment", *Sociological Theory*, **19** (3), 293–311.

OLIVER, P. E., MARWELL, G. and TEIXEIRA, R. (1985), "A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action", *American Journal of Sociology*, **91** (3), 522–556.

OLSON, M. (1968), *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, MA: Harvard University Press).

PALFREY, T. R. and ROSENTHAL, H. (1984), "Participation and the Provision of Discrete Public Goods: A Strategic Analysis", *Journal of Public Economics*, **24** (2), 171–193.

PALFREY, T. R. and ROSENTHAL, H. (1985), "Voter Participation and Strategic Uncertainty", *American Political Science Review*, **79** (1), 62–78.

RAPOPORT, A. (1985), "Provision of Public Goods and the MCS Experimental Paradigm", *American Political Science Review*, **79** (1), 148–155.

RAYMOND, E. S. (1998), "The Cathedral and the Bazaar", *First Monday*, **3** (3), 1–20 (http://www.firstmonday.dk/issues/issue3_3/raymond/).

SHIBATA, H. (1971), "A Bargaining Model of the Pure Theory of Public Expenditure", *Journal of Political Economy*, **79** (1), 1–29.

van de KRAGT, A. J. C., ORBELL, J. M. and DAWES, R. M. (1983), "The Minimal Contributing Set as a Solution to Public Goods Problems", *American Political Science Review*, **77** (1), 112–122.

WARR, P. (1983), "The Private Provision of a Public Good is Independent of the Distribution of Income", *Economics Letters*, **13** (2–3), 207–211.

YOUNG, H. P. (1993), "The Evolution of Conventions", *Econometrica*, **61** (1), 57–84.