Games and Economic Behavior 153 (2025) 474–498

Contents lists available at ScienceDirect

Games and Economic Behavior

journal homepage: www.elsevier.com/locate/geb

Maxmin, coalitions and evolution ☆

Jonathan Newton [*], Miharu Naono [1]

Institute of Economic Research, University of Kyoto, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan

ARTICLE INFO

JEL classification:
C71
C73
D01

Keywords:
Maxmin
Coalitions
Evolution

ABSTRACT

Maxmin decision making can take place at an individual or a coalitional level. We allow evolution to choose between the two, determining the relative shares of individual and coalitional decision making. We consider factors that favor or disfavor the evolution of coalitionality and apply our framework to social dilemmas, oligopolistic price competition and voting on committees.

1. Introduction

This paper connects two important concepts in game theory, maxmin choice and coalitional choice, via the mechanism of evolution. Maxmin chooses strategies that maximize (over one's own strategies) minimum (over the strategies of other players) payoffs. That is, maxmin guarantees the highest payoffs that can be guaranteed without assuming anything about the behavior of other players. Coalitional choice is when players jointly choose strategies to Pareto improve on individual choice. We study the extent to which coalitional choice evolves when decisions are governed by maxmin behavior. Rather than specify the presence or absence of coalitional behavior as an axiom, we allow evolution to choose, partially endogenizing the solution concept that governs how the game is played.

The basic idea is simple. A population comprises collaborative types and non-collaborative types. Players are drawn from the population and matched into groups to play a game. Non-collaborative types play individual maxmin. In contrast, collaborative types look for opportunities to play coalitional maxmin and increase their worst-case payoffs relative to individual maxmin. If the average realized payoff of collaborative types in the population is higher than that of non-collaborative types, then the share of collaborative types in the population will grow. If the reverse is true, then the share of collaborative types in the population will shrink. The equilibrium share of collaborative types depends on the nature of the game and the way in which players are matched to play the game.

By definition, coalitional maxmin increases maxmin payoffs. If this leads to an increase in realized payoffs, then there will be a non-zero share of collaborative types at equilibrium [Theorem 1]. A popular class of games for which this holds are the class of social dilemmas (for example, the prisoner's dilemma), in which individual maxmin leads to low realized payoffs for all players [Lemma 2]. If a game is such that coalitional maxmin requires the participation of every player, then collaborators comprise the entire population

☆ Newton's work was supported by KAKENHI B funding from the Japan Society for the Promotion of Science (Grant 21H00695). We thank Romans Pancs, as well as seminar audiences at the LEG2022 and ICSD2022 conferences.

* Corresponding author.
E-mail addresses: newton@kier.kyoto-u.ac.jp (J. Newton), m.naono2020@gmail.com (M. Naono).
URL: https://jonathannewton.net (J. Newton).
[1] Independent researcher. Work carried out at Institute of Economic Research, University of Kyoto, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan.

https://doi.org/10.1016/j.geb.2025.08.002
Received 28 August 2024

Available online 12 August 2025

at equilibrium [Corollary 1]. Minimum effort public goods games satisfy this condition [Theorem 4]. In contrast, if coalitional maxmin can be conducted by strict subsets of the player set, then there is a trade-off between

1. The benefits of successful coalitional behavior, and
2. Free riding on the coalitional behavior of others.

For example, in n-player prisoner's dilemmas, the share of collaborators is positive and non-monotonic in the cost of contribution, generally decreasing towards zero but making strictly positive jumps at costs where the number of players required for successful collaboration changes [Theorem 2]. This contrasts with threshold public goods games, in which the number of players required for coalitional maxmin is governed by a separate parameter, so there are no such jumps [Theorem 3].

In a model of price competition, we consider two families of demand functions. Under binomial matching, both imperfect substitutes [Theorem 5] and Bertrand competition [Theorem 6] lead to collaborative types taking over the entire population to collude in price setting. In contrast, when demand is a convex combination of these two demand functions, there may be multiple equilibria, and there may exist equilibria with zero, one, or interior shares of collaborative types [Theorem 7]. The intuition is that behavior is predominantly governed by the market with more elastic demand given maxmin conjectures about the behavior of other players. Collusion with respect to this market gives discrete positive externalities to non-colluding firms via the Bertrand market. Sometimes this latter effect dominates and we obtain equilibrium shares of collaborative types that are less than one.

In another application, we consider a committee in which members vote over possible policy changes. Each member has a policy change that they favor, a policy change that they disfavor, and, for a given outcome, a small preference for voting for their favored change. Each member's preferences are independently drawn from an exogenous distribution. By collaborating to coordinate his vote with other committee members, a member can avoid negative outcomes. Therefore, it is perhaps unsurprising that there exists an equilibrium at which collaborative types comprise the whole population. However, there is also the possibility that collaborative members with opposed preferences collaborate to maintain the status quo and avoid their disfavored outcomes. In the absence of collaboration, one of these policies would have been selected. Consequently, if preferences for favored outcomes are stronger than preferences against disfavored outcomes, then there exists an equilibrium in which non-collaborative types comprise the whole population [Theorem 8].

In each application of our model, typical play differs according to the endogenously determined share of collaborative types in the population. Rather than assume that no agent can act a part of a coalition, or assume that every agent can, or assume some fixed probability of any given agent being able to act as part of a coalition, we endogenize this probability. Whether it is stable for there to be many or few collaborative types depends on the game and its parameters. Naturally, there are modeling choices involved in what we do here. We discuss some of them now.

1.1. Modelling choices

Maxmin choice focuses on security in decision making and has been around since the early days of game theory (Von Neumann and Morgenstern, 1947), appearing in diverse areas such as ambiguity aversion (Gilboa and Schmeidler, 1989), preferences for verifiability (Rommeswinkel, 2023), and information design in Bayesian persuasion (Dworczak and Pavan, 2022) and binary-action supermodular games (Morris et al., 2023). An advantage of maxmin is that it is determined independently of what other players do. Consequently, we can talk about maxmin behavior without reference to pre-existing social norms of behavior.

The coalitional version of maxmin that we use is based on the concept of **α-effectiveness** (Aumann and Peleg, 1960). A coalition is α-effective for a payoff vector if it can assure itself of receiving at least that payoff vector, independently of the actions of players outside of the coalition. Coalitional maxmin involves players in a coalition choosing strategies that (i) give higher guaranteed payoffs than individual maxmin, and (ii) cannot lead to a payoff vector that is worse than some payoff vector for which they are α-effective. That is, coalitional maxmin combines an individual rationality condition and an efficiency condition, both with respect to the partial ordering induced by maxmin considerations.

To determine the equilibrium amount of collaborativeness in a population we use the concept of an **evolutionarily stable state** (Taylor and Jonker, 1978). From such a state, if we slightly increase (resp. decrease) the share of collaborative types, then their expected payoff falls below (resp. rises above) the expected payoff of non-collaborative types. Consequently, (regular) evolutionarily stable states are locally asymptotically stable under a broad range of evolutionary dynamics under which traits (e.g. collaborativeness or non-collaborativeness) that lead to relatively high payoffs spread in the population. These include popular dynamics such as the best response and replicator dynamics (Sandholm, 2010; Cressman, 1997).

Aside from the game itself, how players are **matched** can influence the equilibrium share of collaborative types and therefore how the game is typically played. Our general results are proved for balanced matchings, a mild condition to ensure that matching probabilities are measure preserving. This allows for matchings that are positively assortative, negatively assortative or not assortative at all. Perhaps the simplest balanced matching, which we use in several applications, is binomial matching, in which each position in a game is filled by an independent draw from the population, with the probability of a collaborative type being drawn being equal to the share of collaborative types in the population.

1.2. Related literature

Many concepts in game theory are founded upon collaborative choice. Examples include the Core (Gillies, 1959), Strong Equilibrium (Aumann, 1959), the α-core (Aumann and Peleg, 1960), Pairwise Stable Matchings (Gale and Shapley, 1962), Coalition Proofness (Bernheim et al., 1987), Renegotiation Proofness (Farrell and Maskin, 1989), Farsighted Coalitional Stability (Chwe, 1994), Pairwise Stable Networks (Jackson and Watts, 2002), Farsightedly Stable Networks (Herings et al., 2009), Coalitional Rationalizability (Ambrus, 2009), Bayesian Coalitional Rationalizability (Luo and Yang, 2009), Coalitional Stochastic Stability (Newton, 2012), the Farsighted Stable Set (Ray and Vohra, 2015, 2019; Newton, 2021b).

Regarding the evolution of collaborative choice, there is relatively little work. Bacharach (2006) gives a mainly non-quantitative argument as to why a group selection mechanism would lead to collaborative 'team reasoning' in coordination problems and social dilemmas. Angus and Newton (2015) show that this is not true for coordination games on networks, as coalitional behavior can increase the time taken to reach efficient outcomes (Newton and Angus, 2015). In contrast, the current paper considers general games and does not use group selection.

Rusch (2019) conducts a comprehensive study of two strategies, two players (2x2) games. Unlike previous work, players may try and fail to collaborate due to the other player being a different type. Similar to the current paper, the status quo without collaboration is a maxmin strategy profile. In a 2x2 game, it is relatively easy to identify what collaboration will look like. For example, in a prisoner's dilemma, it will involve both players cooperating. For general games, there may be different ways in which players can collaborate, so we should specify which we consider. The current paper complements maxmin at an individual level with maxmin at a coalitional level, defined using the concept of α-effectiveness.

Newton (2017) considers general games under random matching. In contrast to the maxmin approach of the current paper, the status quo without collaboration is an arbitrary Nash equilibrium of the game and collaboration is defined using coalitional better response. The outcome following collaboration will not necessarily be a best response for either individuals or coalitions, which can be seen as antithetic to the fundamentals of the model, which are defined using best/better response. A similar tension does not arise in the current paper, as regardless of the actions of other players, both individual and coalitional maxmin strategies remain maxmin. Another issue is that initializing games at Nash equilibria assumes an unmodeled social choice problem has been solved in the background to establish conventional play. However, if conventional play were routinely subject to collaborative deviations, we would expect that the convention itself would change. Again, this problem does not arise under the maxmin formulation.[2] Finally, we note that Newton (2017) includes discussion of philosophy,[3] developmental psychology[4] and anthropology/primatology[5] vis-à-vis collective agency. We direct the interested reader to that paper and here instead focus on economic applications.

Naturally, the study of the evolution of collaborative choice shares some similarities with the evolution of preferences (e.g. Güth and Kliemt, 1998; Robson, 1996; Samuelson, 2001; Dekel et al., 2007; Heifetz et al., 2007; Frenkel et al., 2018; Heller, 2014; Bergstrom, 1995; Heller and Nehama, 2023). In that literature, it is necessary to specify an outcome of games under particular preferences before considering the evolution of those preferences. Similarly, when studying the evolution of collaborative choice, it is necessary to specify the outcome of games under particular collaborative proclivities before considering the evolution of those proclivities. However, when preferences evolve, individuals' rankings over outcomes change. When collaboration evolves, rankings over outcomes remain the same and it is the constraints on strategy choice that change, in particular the degree of coalitional rationality involved.

A further distinction is with the literature on the evolution of cooperation, where "cooperation" refers to playing the strictly dominated action in a prisoner's dilemma or similar game. That is, cooperation refers to a specific action in a specific game. In contrast, collaboration refers to agency, the 'who' in a strategic situation rather than the 'what'. Collaboration and its evolution can be considered for any game. The question of when collaboration can aid cooperation has been considered in a network setting in Angus and Newton (2020). In general, however, collaboration is completely independent of cooperative or altruistic behavior. Indeed, collaboration opens up the possibility of coalitions of players ganging up to inflict suffering on others. In Section 3.1 we will see the interaction between collaboration and cooperation in the classic prisoner's dilemma setting.

2. Model and fundamental results

Individuals from a **population** are matched into groups to play a **game**. How the game is played depends on the **types** of the individuals. The **fitness** of a type is given by its average payoffs in the game. Given these fitnesses, we consider **evolutionarily stable** shares of each type in the population.

[2] The evolution of conventions is an interesting topic, but not directly relevant to the current paper given modeling choices. The interested reader is directed to Newton (2021a) and citations therein.

[3] E.g. The definition of a collective intention (Tuomela and Miller, 1988; Searle, 1990; Bratman, 1992) and whether they can be reduced to individual intentions (Gold and Sugden, 2007; Butterfill, 2012; Gilbert, 1990; Velleman, 1997).

[4] E.g. The urge to collaborate develops earlier in infancy than does sophisticated logical inference or the ability to articulate hierarchical beliefs (Tomasello and Rakoczy, 2003).

[5] E.g. Humans collaborate more than the other great apes (Tomasello and Herrmann, 2010; Tomasello and Carpenter, 2007) and it is hypothesized that this was crucial to our development of higher intelligence (Call, 2009) and language (Fitch, 2010).

2.1. The game

Interactions are modeled by an n-player normal form game $\Gamma = (N, (S_i)_{i\in N}, (\pi_i)_{i\in N})$, with player set $N = \{1, \dots, n\}$, strategy sets S_i that are closed bounded subsets of Euclidean space, continuous payoff functions $\pi_i(.) : S \to \mathbb{R}$, where $S = \bigtimes_{i\in N} S_i$. That is, for strategy profile $s = (s_1, \dots, s_n) \in S$, the payoff of player i is $\pi_i(s)$.

Outcomes of Γ will involve maxmin play by both individuals and coalitions. We proceed to define the relevant concepts and notation.

Coalitional notation. For $T \subseteq N$, denote $S_T = \bigtimes_{i\in T} S_i$. That is, $s_T \in S_T$ is a strategy profile for coalition T. Denote the vector of payoffs for players in coalition T by $\pi_T(\cdot) = (\pi_i(\cdot))_{i\in T}$. As is standard, let $-i$ and $-T$ denote the sets $N \setminus \{i\}$ and $N \setminus T$ respectively.

Vector inequalities. For $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, $n \geq 2$, let $x > y$ denote $x_i > y_i$ for all $i = 1, \dots, n$; $x \geqq y$ denote $x_i \geq y_i$ for all $i = 1, \dots, n$; and $x \geq y$ denote $x \geqq y$ and $x \neq y$.

Individual maxmin. Denote maxmin strategies and payoffs for player i by

$$\underline{S}_i = \operatorname*{argmax}_{s_i \in S_i} \min_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}), \qquad \underline{\pi}_i = \max_{s_i \in S_i} \min_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}). \tag{1}$$

Compactness of S_i, S_{-i} and continuity of π_i ensure that $\underline{S}_i$ is nonempty and $\underline{\pi}_i$ is well-defined. Let $\underline{S} = \bigtimes_{i\in N} \underline{S}_i$ and $\underline{S}_T = \bigtimes_{i\in T} \underline{S}_i$ denote product sets of maxmin strategies. Let $\underline{\pi} = (\underline{\pi}_i)_{i\in N}$ and $\underline{\pi}_T = (\underline{\pi}_i)_{i\in T}$ denote vectors of maxmin payoffs. Note that vectors of maxmin payoffs are not the same as the payoffs realized when players play maxmin strategies. That is, $\underline{\pi}_T = (\underline{\pi}_i)_{i\in T}$ is not necessarily equal to $\pi_T(\underline{s})$. This distinction is important in what follows.

α-effectiveness. The set of possible payoffs for coalition T when it plays s_T is

$$\Pi_T(s_T) = \left\{ \pi_T(s_T, s_{-T}) : s_{-T} \in S_{-T} \right\}. \tag{2}$$

Coalition T is α-effective (Aumann & Peleg, 1960) for the payoff vector z if it can assure itself of receiving at least z, independently of the actions of players in $N \setminus T$. We shall use a strict version of α-effectiveness under which T can assure itself of a payoff vector better than z, independently of the actions of players in $N \setminus T$.

- T is *α-effective* for $z \in \mathbb{R}^T$ if $\exists\, s_T$: for all $\pi_T \in \Pi_T(s_T)$, $\pi_T \geqq z$.
- T is *strictly α-effective* for $z \in \mathbb{R}^T$ if $\exists\, s_T$: for all $\pi_T \in \Pi_T(s_T)$, $\pi_T \geq z$.

2.2. Coalitional maxmin

Coalition T will be interested in s_T that (i) give higher guaranteed payoffs than individual maxmin, and (ii) are undominated in a maxmin sense. Strategy s_T is dominated if there exists s'_T that guarantees T a payoff vector better than some payoff vector that is possible under s_T. The existence of such an s'_T is equivalent to T being strictly α-effective for some $\pi_T \in \Pi(s_T)$. Thus, we define the set of coalitional maxmin strategies for coalition T,

$$MM(T) = \left\{ s_T \in S_T : \begin{array}{l} \forall \pi_T \in \Pi_T(s_T), \text{ we have } \pi_T \geq \underline{\pi}_T \text{ and} \\ T \text{ is not strictly } \alpha\text{-effective for } \pi_T. \end{array} \right\} \tag{3}$$

Assume that the set of maxmin profiles is nonempty for some possible coalition.

Assumption 1. For some $T \subseteq N$, $|T| \geq 2$, $MM(T) \neq \emptyset$.

The following lemma shows that once coalitional maxmin becomes viable for some coalition T, it remains viable for supersets $T' \supset T$.

Lemma 1. *If $T \subset T'$, $MM(T) \neq \emptyset$, then $MM(T') \neq \emptyset$.*

The following condition holds when, compared to individual maxmin profiles, a coalition T that plays coalitional maxmin *realizes gains* in payoffs.

(RG) For all $T \subseteq N$, if $\underline{s}, \underline{s}' \in \underline{S}$ and $s_T \in MM(T)$, then $\pi_T(s_T, \underline{s}_{-T}) \geq \pi_T(\underline{s}')$.

(RG) holds when coalitional maxmin leads to actual gains in payoffs. It indicates complementarity between acting to secure payoffs and payoffs themselves. It often holds, but it is instructive to consider an example where it does not.

Example 1. Let $N = \{1, 2, 3\}$. For all $i \in N$, let $S_i = \{A, B\}$ and

$$\pi_i(A, s_{-i}) = 5\,|\{j \in N \setminus \{i\} : s_j = A\}|,$$
$$\pi_i(B, s_{-i}) = 2\,|\{j \in N : s_j = B\}|.$$

It follows that $\underline{s} = (B, B, B) \in \underline{S}$, giving $\underline{\pi} = (\underline{\pi}_1, \underline{\pi}_2, \underline{\pi}_3) = (2, 2, 2)$. Observe that this differs from the realized payoffs from maxmin strategies of $\pi_i(\underline{s}) = 6$ for all $i \in N$. Further, for $T = \{i, j\} \subset N$, $s_T = (A, A) \in MM(T)$ guarantees a payoff of $(5, 5)$ for T. However, in terms of realized payoffs, $\pi_T(s_T, \underline{s}_{-T}) = \pi_T((A, A), B) = (5, 5)$, whereas $\pi_T(\underline{s}) = \pi_T((B, B, B)) = (6, 6)$, so (RG) is not satisfied.[6]

2.3. Types and behavior

An individual playing a game may be in a *collaborative mood*. The coalition of individuals in a collaborative mood will play coalitional maxmin whenever the set of coalitional maxmin strategies is nonempty. We refer to individuals in a collaborative mood as α-types. We choose this name because coalitional maxmin is determined by the constraints of α-effectiveness. Individuals who are not in a collaborative mood play individual maxmin. We refer to such individuals as ν-types.

Consider a group N of n individuals who encounter a game Γ. Let $N_\alpha \subseteq N$ denote the set of α-type individuals within the group. The outcome of the game will be a strategy profile s^* satisfying:

(C) 1. If $MM(N_\alpha) \neq \emptyset$, let s^* be chosen according to some probability measure $G_{N_\alpha,\Gamma}(.)$ on $\left\{ s \in S : s_{N_\alpha} \in MM(N_\alpha),\ s_{-N_\alpha} \in \underline{S}_{-N_\alpha} \right\}$.
2. If $MM(N_\alpha) = \emptyset$, let s^* be chosen according to some probability measure $G_\Gamma(.)$ on $\underline{S}$.

Without specifying $G_{N_\alpha,\Gamma}(.)$ and $G_\Gamma(.)$, condition (C) is not a complete description of behavior. In particular, when there are multiple coalitional maxmin profiles, (C-i) specifies that one such profile will be chosen by N_α, but allows it to be any profile in $MM(N_\alpha)$ with any probability. (C-ii) implies that when there is no coalitional maxmin available, α-types and ν-types are indistinguishable.

2.4. Matching

Consider a population comprising unit mass of individuals, each of whom may be of α or ν-type. Let the share of α-types in the population be x and the share of ν-types be $1 - x$.

Each member of the population is matched to play Γ in a group of n individuals. We assume that the allocation of individuals to player positions in the game Γ is independent of type.[7] Given a population state x, and an individual, let Z be a random variable denoting the number of α-types amongst the other $n-1$ individuals with whom the given individual is matched. A *matching protocol* specifies Z for all x and all values of k from 0 to $n-1$.

Binomial matching. Perhaps the simplest matching protocol is binomial matching, where group members are drawn uniformly and independently from the population, so that the probability of any given group member being α-type is x, regardless of the types of other group members. This gives

$$Pr_x[Z = k] = \binom{n-1}{k} x^k (1-x)^{n-1-k}. \tag{4}$$

Assortativity. The model can also handle *assortative* matching probabilities that depend on type. Write $Pr_x[Z = k \mid \alpha]$ and $Pr_x[Z = k \mid \nu]$ as the probabilities that there are k α-type individuals amongst the other members of the group, conditional on a given individual being α-type and ν-type respectively. Assume $Pr_x[Z = k \mid \alpha]$ and $Pr_x[Z = k \mid \nu]$ are continuous in x, and strictly positive for $x \in (0, 1)$.

Balanced matchings. Any α-type has a probability $Pr_x[Z = k-1 \mid \alpha]$ of being in a group that includes exactly k α-types, including himself. Therefore, the mass of α-types in such groups equals $x\, Pr_x[Z = k-1 \mid \alpha]$. Any ν-type has a probability $Pr_x[Z = k \mid \nu]$ of being in a group that includes exactly k α-types. Therefore, the mass of ν-types in such groups equals $(1-x)\, Pr_x[Z = k \mid \nu]$. Now, any group with k α-types has $n-k$ ν-types, so the ratio of α-type individuals in such groups to ν-type individuals in such groups must equal $k/(n-k)$. We have

$$\frac{x\, Pr_x[Z = k-1 \mid \alpha]}{(1-x)\, Pr_x[Z = k \mid \nu]} = \frac{k}{n-k} \quad \text{for} \quad k = 1, \ldots, n-1. \tag{5}$$

We assume that the matchings we consider are balanced in this way. It can be checked that binomial matching satisfies this condition.

2.5. Evolutionary stability

Given a game Γ, some behavioral rule satisfying (C), and some matching protocol, let $f_\alpha(x)$, $f_\nu(x)$ denote the fitnesses, that is the expected payoffs, of α and ν-types respectively at population state x. Assume that $G_{N_\alpha,\Gamma}(.)$ and $G_\Gamma(.)$ are such that these expectations exist.

In Appendix A, we see that $f_\alpha(x)$ and $f_\nu(x)$ are continuous in x, therefore the definition of a evolutionarily stable state (Taylor and Jonker, 1978) simplifies.

[6] Note that when (RG) does not hold, this does not make coalitional maxmin irrational. In this example, by playing (A, A), coalition T guarantees payoffs $(5, 5)$, whereas any other strategies chosen by T would carry the possibility of lower payoffs.

[7] If allocation to player positions depended on type, then in asymmetric games we could give α or ν-types an advantage by disproportionately allocating them to high payoff positions. Besides being somewhat trivial, this is not in the spirit of our study, which wishes to consider the behavioral implications of α or ν-types. A more interesting extension would be model situations in which different positions in the game are associated with different type distributions by introducing multiple populations, with each population associated with one or more positions in the game.

Evolutionarily stable state (ESS). An interior state $x^* \in (0,1)$ is an ESS if and only if $f_\alpha(.) - f_\nu(.)$ is strictly decreasing in some neighborhood of x^* and equal to 0 at x^*. The extremal state $x^* = 0$ ($x^* = 1$) is an ESS if and only if $f_\alpha(.) - f_\nu(.)$ is strictly negative (positive) in some open interval bounded below (above) by x^*.

An ESS is a state such that, following the invasion of the population by a small share of mutants, the non-mutant share of the population outperforms the invading mutants. Such states exhibit stability properties under popular dynamics such as the best response and replicator dynamics (Sandholm, 2010; Cressman, 1997).[8]

The replicator dynamic, used extensively in evolutionary biology and the social sciences, models the rate of growth of a trait as proportional to the fitness of holders of the trait relative to average fitness. This typically represents high payoffs leading to greater biological reproduction or to a higher probability of the trait being imitated by other agents. In our model, it is a simple way of thinking about how the shares of α and ν-types might adjust when high payoffs are associated with reproductive success. Under the replicator dynamic, the share of α-types grows at a rate proportional to the fitness advantage of α-types relative to average fitness,

$$\frac{\dot{x}}{x} = f_\alpha(x) - \underbrace{\big(x f_\alpha(x) + (1-x) f_\nu(x)\big)}_{\text{average fitness}} = (1-x)(f_\alpha(x) - f_\nu(x)). \tag{6}$$

2.6. Proliferation of α-types

It turns out that, for any game satisfying (RG) and any matching protocol, if α-types within groups collaborate to play coalitional maxmin (C), then α-types will make up a positive share of the population at any ESS. That is, the state $x = 0$ is always vulnerable to invasion by α-types, no matter how small the mass of invaders.

Theorem 1. *If (C), (RG) hold, then $x > 0$ in any ESS.*

The evolution of collaboration is non-trivial even when (RG) is satisfied because of the possibility of free riding by non-collaborative types. For example, consider a three player threshold public goods game, where each player can contribute or not, with contribution having a cost of $c > 0$. The public good, worth $b > c$ to every player, is provided if and only if at least two players contribute. Individual maxmin is to not contribute, giving maxmin payoffs of 0 for each player. Consider two α-types and one ν-type who are matched to play this game. The α-types play coalitional maxmin, both contribute and obtain payoff $b - c > 0$. The ν-type plays individual maxmin, does not contribute and obtains a higher payoff of b. In this way, the ν-types can free ride on the collaborative behavior of α-types. An important question is then whether this free riding effect can be strong enough to give an ESS at $x = 0$. Theorem 1 answers this in the negative.

Theorem 1 is broad in that it applies to many games. (RG) only requires that when a coalition assures its members of higher payoffs, this leads to higher realized payoffs. We shall show that (RG) is satisfied by prisoner's dilemmas, threshold public goods games, minimum effort games, and price competition with imperfect substitutes, the latter under a restriction to symmetric elements of $MM(N_\alpha)$.

Theorem 1 tells us that there is no ESS at $x = 0$, but leaves open the possibility of an ESS close to zero. Indeed, in Section 3.2 we show that the threshold public goods game has a unique ESS which approaches zero as $b/c \to 1$ and approaches one as $b/c \to \infty$. In this sense, Theorem 1 on its own is weak. Stronger results are derived for specific applications later in the paper (Sections 3-5) and in Appendix B we discuss the inverse of Theorem 1, conditions under which $x = 0$ is an ESS.

Summary of Theorem proof. The proof can be summarized in three steps. Firstly, coalitional maxmin (C) implies that α-types increase their guaranteed payoffs relative to what they can guarantee by individual maxmin. Secondly, (RG) tells us that when α-types play coalitional maxmin and guarantee higher payoffs than those guaranteed by individual maxmin, they also improve their realized payoff relative to their realized payoff from individual maxmin. Finally, the balance condition on matchings implies that when α-types are a small share of the population, any given α-type will find himself in a group in which coalitional maxmin occurs much more frequently than any given ν-type finds himself in such a group. Thus an α-type will enjoy the benefits of collaboration far more often than an ν-type will get the opportunity to free ride on the collaboration of others.

Next, we consider a special, but important case. Note that when $MM(T) \neq \emptyset \Leftrightarrow T = N$, then coalitional maxmin will always involve every group member. As ν-types will never be members of a group in which collaboration occurs, α-types outperform ν-types regardless of their share of the population.

Corollary 1. *If (C), (RG) hold and $MM(T) \neq \emptyset \Leftrightarrow T = N$, then there exists a unique ESS $x^* = 1$.*

In general, for $x = 1$ to be an ESS we require small invasions of ν-types to underperform incumbent α-types. Under (RG), this corresponds to gains from free riding by invading ν-types being smaller than average realized gains of α-types from coalitional

[8] For extremal ESS, say $x^* = 0$, we require $f_\alpha(x) - f_\nu(x) < 0$ for $x \in (0, \varepsilon)$ for some $\varepsilon > 0$, but allow $f_\alpha(0) - f_\nu(0) = 0$. The ESS may not be *regular* (Taylor and Jonker, 1978), so asymptotic stability results of Sandholm (2010); Cressman (1997) do not directly apply. However, as long as dynamics respect fitness differences, x will decrease from any starting point in $(0, \varepsilon)$. Indeed, $\Omega_{x^*=0}(x) = x$ is a trivial local Lyapunov function for $x^* = 0$ under such dynamics in our setting.

maxmin. Corollary 1 is a special case of this, as when collaboration requires the participation of all players, there are no gains from free riding.

We now move to applications of our model to specific classes of games. Aside from the independent interest of these applications, we shall refer to them in Appendix B when we discuss robustness of and extensions to our modeling framework.

3. Social dilemmas

The following condition, satisfied by social dilemmas such as the prisoner's dilemma and stag hunt, specifies that when every player plays individual maxmin, they obtain maxmin payoffs.

(SD) If $s \in \underline{S}$, then $\pi(s) \in \underline{\pi}$.

Lemma 2. *(SD) implies (RG).*

Therefore, social dilemmas satisfy the realized gains property, so that Theorem 1 applies and there are a positive share of α-types at any ESS. For given examples, we shall make more detailed statements.

3.1. n-player prisoner's dilemma

For all $i \in N$, let $S_i = \{0, 1\}$, where 1 represents contribution towards a public good and 0 represents non-contribution. The payoff for $i \in N$ is given by

$$\pi_i(s) = b \sum_{j \in N} s_j - c\, s_i, \tag{7}$$

where b represents a public benefit from a player's contribution, c is the cost of contribution, and $nb > c > b$. Let m be the least integer such that $mb > c$.

The set of individual maxmin strategies for player i is $\underline{S}_i = \{0\}$. For coalitional maxmin, considering $T \subseteq N$,

$$\begin{aligned} &\text{If} \quad |T| < m, \quad \text{then} \quad MM(T) = \varnothing. \\ &\text{If} \quad |T| \geq m, \quad \text{then} \quad MM(T) = \left\{ s_T : \sum_{i \in T} s_i > \max\{m-1\,,\, |T| - m\} \right\}. \end{aligned} \tag{8}$$

That is, a coalitional maxmin for T involves contribution by at least m players in T, with strictly fewer than m players not contributing.

As the game satisfies (SD), Lemma 2 tells us that (RG) is satisfied. Thus, for any behavioral rule satisfying (C) and any matching protocol, Theorem 1 tells us that α-types comprise a strictly positive share of the population at any ESS. Furthermore, if $m = n$, then Corollary 1 tells us that the unique ESS is for α-types to comprise the entire population.

For more precise results, we specify a matching protocol and behavioral rule. For the matching protocol we consider binomial matching. For the behavioral rule, we focus on the coalitional maxmin in which every α-type contributes,

$$\begin{aligned} &\text{For} \quad |N_\alpha| < m, \qquad G_\Gamma(\underbrace{0, \ldots, 0}_{\in \underline{S}}) = 1. \\ &\text{For} \quad |N_\alpha| \geq m, \qquad G_{N_\alpha, \Gamma}(\underbrace{1, \ldots, 1}_{\in MM(N_\alpha)}, \underbrace{0, \ldots, 0}_{\in \underline{S}_{-N_\alpha}}) = 1. \end{aligned} \tag{9}$$

Under the above assumptions, we have the following results.

Theorem 2. *In the n-player prisoner's dilemma, when matching is binomial and coalitional maxmin involves contribution by all α-types, we have*

1. *There is a unique ESS, x^*. If $m < n$, then $x^* \in (0, 1)$. If $m = n$, then $x^* = 1$.*
2. *From any mixed population such that $x \in (0, 1)$, the replicator dynamic converges to x^*.*
3. *x^* strictly decreases in n, is non-monotonic in c for given b.*
4. *$x^* \to 1$ as $c \downarrow b$.*
5. *For $p \in \mathbb{N}$, $2 \leq p \leq n-1$, $x^* \to 0$ as $c \uparrow pb$. $x^* > 0$ if $c = pb$.*
6. *$x^* = 1$ if $c \geq (n-1)b$.*

Theorem 2 is illustrated in Fig. 1. For given b, as c increases, the share of α-types at ESS decreases as the payoff benefits from the marginal collaboration approach zero. Eventually a threshold is reached at which the number of players required for a mutually profitable collaboration increases. At such a point, the share of α-types at ESS jumps discontinuously upwards, before proceeding to

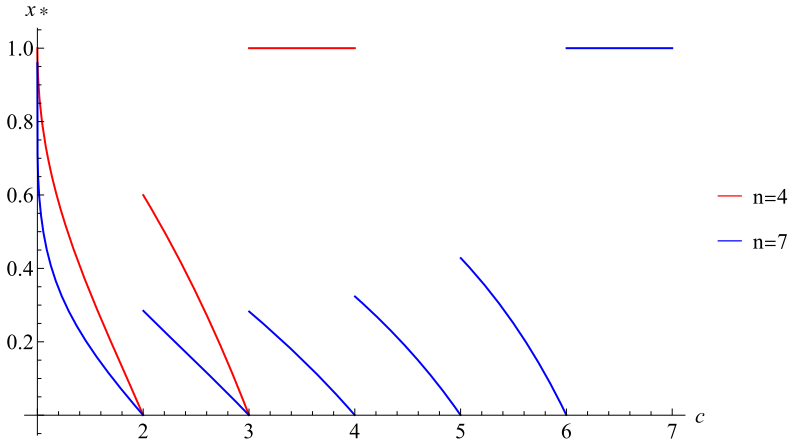

Fig. 1. *n*-player prisoner's dilemmas. Fixing the public benefit (per player) from each contribution at $b = 1$, the ESS share x^* of collaborative α-types changes with the cost of contribution c for different numbers of players n. The ESS share of collaborators is positive and non-monotonic in the cost of contribution, decreasing towards zero but making strictly positive jumps at costs where the number of players required for successful collaboration changes [Theorem 2]. If coalitional maxmin requires the participation of every player, then $x^* = 1$ [Corollary 1]. For higher values of c, specifically $c \geq n$, the game is no longer a prisoner's dilemma and there exists no coalitional maxmin, thus no difference in behavior or fitness across types.

decrease again. Eventually, c becomes large enough that the number of players necessary for a profitable collaboration equals n, so Corollary 1 implies that the share of α-types at ESS must equal one.

3.2. Threshold public goods game

For all $i \in N$, let $S_i = \{0, 1\}$, where 1 represents contribution towards a public good and 0 represents non-contribution. The payoff for $i \in N$ is given by

$$\pi_i(s) = \begin{cases} b - cs_i & \text{if } \sum_{j \in N} s_j \geq m \\ -cs_i & \text{otherwise} \end{cases}. \tag{10}$$

where b represents a public benefit that is only provided if at least $m \in \mathbb{N}$ players contribute, c is the cost of contribution, and $b > c > 0$.

The set of individual maxmin strategies for player i is $\underline{S}_i = \{0\}$. For coalitional maxmin, considering $T \subseteq N$, a coalitional maxmin involves contribution by exactly m players in T.

$$\text{If} \quad |T| < m, \quad \text{then} \quad MM(T) = \emptyset. \tag{11}$$

$$\text{If} \quad |T| \geq m, \quad \text{then} \quad MM(T) = \left\{ s_T : \sum_{i \in T} s_i = m \right\}.$$

As the game satisfies (SD), Lemma 2 tells us that (RG) is satisfied. Thus, for any behavioral rule satisfying (C) and any matching protocol, Theorem 1 tells us that α-types comprise strictly positive share of the population at any ESS. Furthermore, if $m = n$, then Corollary 1 tells us that the unique ESS is for α-types to comprise the entire population.

For more precise results, we again use binomial matching as our matching protocol. For our behavioral rule, we do not have to give more details. It is possible that $G_{N_\alpha,\Gamma}(\cdot)$ might favor contribution by some player positions over others, but even if this is the case, ex-ante, the probability of being in such a position conditional on being an α-type will be the same.

Under the above assumptions, we have the following results.

Theorem 3. *In the threshold public goods game, when matching is binomial and (C) holds, we have*

1. *There is a unique ESS, x^*. If $m < n$, then $x^* \in (0, 1)$. If $m = n$, then $x^* = 1$.*
2. *From any mixed population such that $x \in (0, 1)$, the replicator dynamic converges to x^*.*
3. *x^* strictly decreases in n, increases in $\frac{b}{c}$.*
4. *If $m < n$, then $x^* \to 0$ as $\frac{b}{c} \to 1$ and $x^* \to 1$ as $\frac{b}{c} \to \infty$. In particular, x^* may be greater or less than $\frac{m}{n}$.*

Theorem 3 is illustrated in Fig. 2. For $m < n$, the share of α-types at ESS approaches one as c approaches zero and approaches zero as c approaches b. In contrast, for $m = n$, the share of α-types at ESS equals one for all $c < 1$.

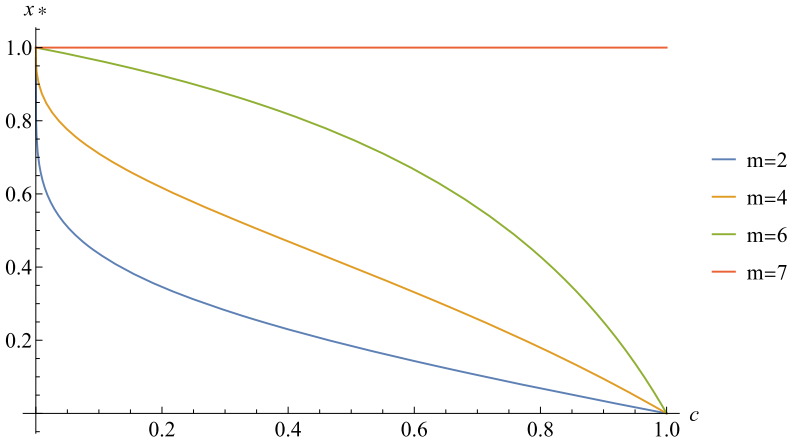

Fig. 2. Threshold public goods. Fixing the number of players $n = 7$ and the benefit (per player) from the public good $b = 1$, the ESS share x^* of collaborative α-types changes with the cost of participation in public goods provision c for different thresholds m. For $m < n$, the ESS share of collaborators is positive, approaching 1 as $c \to 0$ and approaching 0 as $c \to 1$ [Theorem 3]. If $m = n$, then coalitional maxmin requires the participation of every player, so $x^* = 1$ [Corollary 1].

3.3. Minimum effort game

For all $i \in N$, let $S_i = [0, \bar{e}] \subset \mathbb{R}$, where $s_i \in S_i$ is effort towards accomplishing some task, with $\bar{e}$ the maximum possible effort. The payoff for $i \in N$ is given by

$$\pi_i(s) = b \min_{j \in N} s_j - c\, s_i, \tag{12}$$

where b represents a public benefit that scales with the minimum effort across all players, c is the marginal cost of effort, and $b > c > 0$.

The set of individual maxmin strategies for player i is $\underline{S}_i = \{0\}$. For coalitional maxmin, considering $T \subseteq N$,

$$\text{If} \quad T \neq N, \quad \text{then} \quad MM(T) = \varnothing. \tag{13}$$

$$\text{If} \quad T = N, \quad \text{then} \quad MM(T) = \left\{ s_T : \forall i \in T, s_i = \bar{e} \right\}.$$

As the game satisfies (SD), Lemma 2 tells us that (RG) is satisfied. Furthermore, (13) tells us that a coalitional maxmin requires every player to participate. Thus, for any behavioral rule satisfying (C) and any matching protocol, Corollary 1 tells us that the unique ESS is for α-types to comprise the entire population.

Theorem 4. *In the minimum effort game, for any matching protocol, when (C) holds we have a unique ESS* $x^* = 1$.

4. Price competition

Consider price competition by a set of firms N. A strategy for each firm is a price $s_i \in S_i = [0, P]$. Demand for the good produced by firm i is given by

$$D_i(s) = \left[D - a\, s_i + b \sum_{j=1}^{n} s_j \right]_+, \qquad D, a, b > 0 \tag{14}$$

Assume $a > bn$, so a firm's own price has a larger effect on demand for its product than the prices set by other firms. Let $P > {}^{D}/_{2(a-bn)}$ so that P will not bind for the remainder of our analysis. Assume that firms can produce at zero cost, so that payoffs are

$$\pi_i(s) = s_i\, D_i(s). \tag{15}$$

The demand for the product of firm i increases in the prices set by the other firms. Hence, the payoffs of firm i are always lowest when the other firms all set prices equal to zero. Given this, we obtain that the unique maxmin strategy is

$$\underline{s}_i = \frac{D}{2(a-b)}. \tag{16}$$

There are many coalitional maxmin strategies for $T \subseteq N$, $|T| \geq 2$. Consider those in which every player $i \in T$ sets the same price s_α. Again, the worst case is for all $j \notin T$ to set $s_j = 0$. The payoff for each firm in T is

$$s_\alpha \left(D - a\, s_\alpha + b\, |T|\, s_\alpha \right). \tag{17}$$

Maximizing with respect to s_α, we obtain

$$s_\alpha = \frac{D}{2(a - b|T|)}. \tag{18}$$

Increased prices from coalitional maxmin by α-types exert positive externalities on ν-types by increasing demand. This raises the possibility that these externalities outweigh any positive payoff benefits for α-types. Naturally, this depends on both payoffs and the matching rule.

Theorem 5. *In the price competition game with demand given by (14), when coalitional maxmin involves all firms in a coalition charging the same price,*

1. $x > 0$ *in any ESS.*
2. *Under binomial matching, there is a unique ESS* $x^* = 1$.

4.1. Bertrand competition

Consider an alternative demand function that shares demand equally amongst the firms that charge the lowest price, with firms charging a higher price receiving no demand. For maximum demand D', $P > D'/2$, we have

$$D'_i(s) = \begin{cases} \frac{[D' - s_i]_+}{|\operatorname{argmin}_{i \in N} s_i|} & \text{if } i \in \operatorname{argmin}_{i \in N} s_i, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

Payoffs are as before, but with an added penalty for having zero customers.

$$\pi_i(s) = s_i\, D'_i(s) - \left(1 - \operatorname{sgn}(D'_i(s))\right)\epsilon, \tag{20}$$

where sgn is the sign function that takes value 1 when its argument is strictly positive and value 0 when its argument is 0, $\epsilon > 0$ is a penalty for having zero demand.

The role of the penalty ϵ is as a tiebreaker in determining maxmin strategies. Specifically, it guarantees a unique maxmin strategy $\underline{s}_i = 0$ associated with the maxmin payoff $\underline{\pi}_i = 0$.

Consider a coalition T of size $|T| < n$. Due to the possibility of some $j \notin T$ playing $s_j = 0$, the set of coalitional maxmin strategies $MM(T) = \emptyset$. If $|T| = n$, then there is a unique coalitional maxmin strategy at which every player chooses the monopoly price,

$$s_\alpha = \frac{D'}{2}. \tag{21}$$

Moreover, players who participate in a coalitional maxmin of size n obtain strictly higher payoffs as a consequence. That is, (RG) holds. Thus, for any behavioral rule satisfying (C) and any matching protocol, Corollary 1 tells us that the unique ESS is for α-types to comprise the entire population.

Theorem 6. *In the Bertrand price competition game given by (19)-(20), for any matching protocol, when (C) holds we have a unique ESS* $x^* = 1$.

4.2. Hybrid demand function

Now consider a demand function that is a convex combination of the demand functions in (14) and (19),

$$D^*_i(s) = \theta\, D_i(s) + (1 - \theta)\, D'_i(s), \qquad \theta \in (0, 1), \tag{22}$$

Payoffs can be given by either of the payoff functions (15) or (20). The analysis is identical as zero demand is not binding when it comes to determining maxmin strategies. Again, the worst outcome for firm i is always when other firms set prices of zero. Given this, individual maxmin strategies $\underline{s}_i$ are the same as in our initial case of demand given by D_i. That is, as given by (16).

A similar argument applies to coalitional maxmin strategies (with equal prices) for T, $|T| < n$, which are given by (18). For $|T| = n$, coalitional maxmin strategies (with equal prices) solve the monopoly problem, and we obtain

$$s_\alpha = \frac{\theta\, D + (1 - \theta)\, D'}{2\theta(a - bn) + 2(1 - \theta)}. \tag{23}$$

Interestingly, despite demand being a convex combination of the two previous demand functions, results differ markedly. In particular, even under Binomial matching there may be an ESS at $x = 0$, there may be multiple ESS, and there will not necessarily be an ESS at $x = 1$.

Theorem 7. *In the price competition game with demand given by (22), under binomial matching, when coalitional maxmin involves all firms in a coalition charging the same price,*

1. *Keeping other parameters fixed and letting* $\theta \to 0$, *eventually*
   (a) $x = 0$ *is an ESS.*

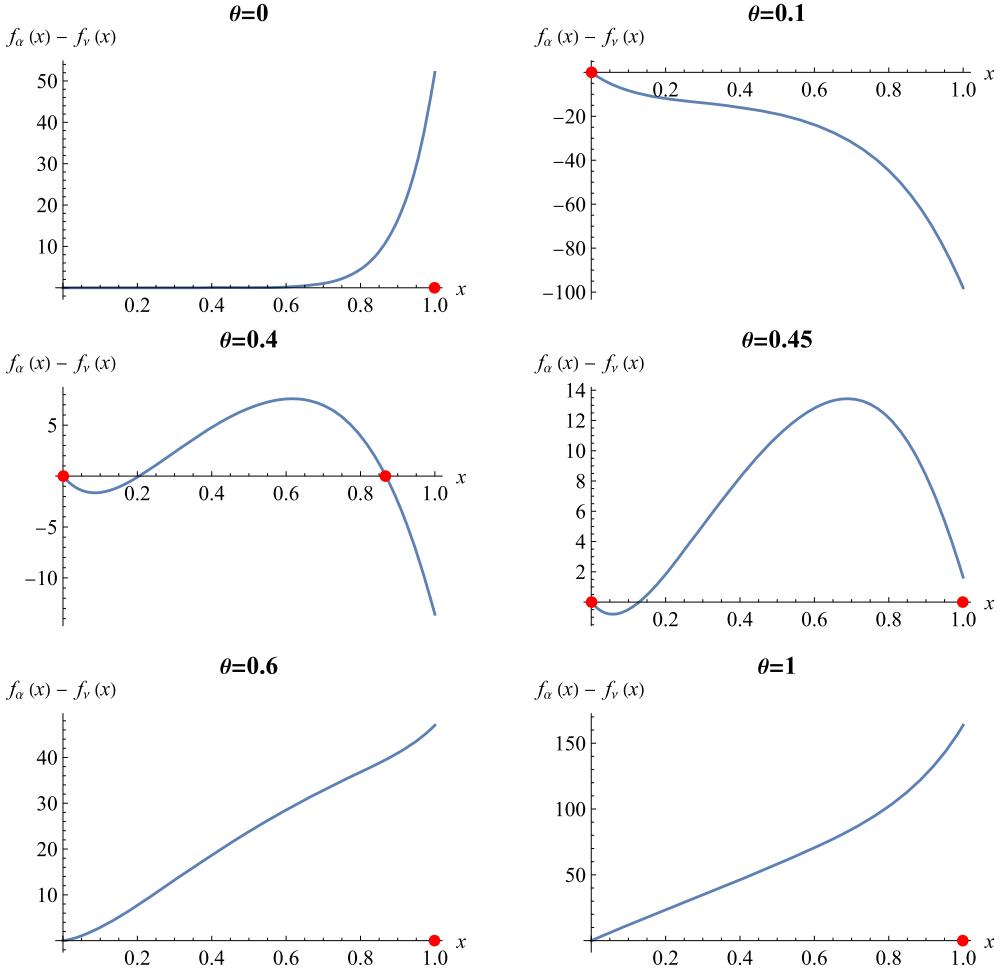

Fig. 3. Price competition. Fixing parameters $n = 12$ $a = 15$, $b = 1$, $D = 100$, $D' = 50$, for different weightings θ of the two markets, the fitness difference between α and ν-types is shown as a function of the share x of α-types in the population. ESS shares are shown as red dots. For $\theta = 0$, $x = 1$ is the unique ESS [Theorem 6]. However, for low positive θ, there is an ESS at $x = 0$, there may be multiple ESS, and $x = 1$ is not necessarily an ESS [Theorem 7(i)]. For high θ, the situation is the same as for $\theta = 1$, with $x = 1$ the unique ESS [Theorems 5, 7(ii)].

(b) *if $D \ll D'$, then $x = 1$ is an ESS.*

(c) *If monopoly prices are identical across markets, $D/2(a-b) = D'/2$, then $x = 1$ is not an ESS.*

2. *Keeping other parameters fixed and letting $\theta \to 1$, eventually there is a unique ESS $x^* = 1$.*

Consider a population of ν-types that is invaded by a small share of α-types. These α-types charge a higher price, which gives them a higher profit in the original market but means they obtain no demand and get no profit from the Bertrand market. When θ is close to zero, the latter effect dominates, so that the α-types are driven to extinction and there is an ESS at $x = 0$ [Theorem 7(i-a)]. Conversely, when θ is close to one, the former effect dominates and the share of α-types grows in the population, eventually reaching $x = 1$ [Theorem 7(ii)].

Coalitional maxmin players only derive profit from the Bertrand market when their coalition has size n. However, as θ goes to zero, this becomes the market that matters for profits. Hence, whether or not $x = 1$ is an ESS depends on whether the profits of a size n coalition are greater than n times the profits of a single ν-type who is matched with $n - 1$ α-types and serves the entire Bertrand market. In extreme cases, for example $D' \gg D$, this holds and there is an ESS at $x = 1$ [Theorem 7(i-b)]. In more moderate circumstances, this is not the case [Theorem 7(i-c)].

Under our hybrid demand function, higher prices due to coalitional maxmin by T, $|T| < n$, increase payoffs of players outside of T by exerting a positive externality on both components of the demand function. The Bertrand component is particularly dramatic in this respect, with ν-types lucky enough to be matched into groups with few ν-types receiving a large boost to their Bertrand demand component. Of course, under most intuitively plausible matching protocols, the probability of being in such a group will rise with the share of α-types in the population, raising the possibility of interior ESS, such as those in Fig. 3.

5. Incomplete information

We shall allow the game to be a game of incomplete information over payoffs. Consider a probability distribution over some set of games $\mathcal{G}$, where each $\Gamma = (N, (S_i)_{i\in N}, (\pi_i)_{i\in N}) \in \mathcal{G}$ has the same player set N and strategies $(S_i)_{i\in N}$, differing only in payoffs $(\pi_i)_{i\in N}$.

Given the types (α or ν) of players in N, after a game $\Gamma \in \mathcal{G}$ is realized, play proceeds exactly as it would if Γ were played with certainty. This is because maxmin play does not depend on the payoffs of other players. However, the probability distribution over $\mathcal{G}$ affects average payoffs (i.e. the fitness) of α-types and ν-types and thus ESS shares of α-types. Consequently, average play across instances of a given $\Gamma \in \mathcal{G}$ is affected by the probability distribution over $\mathcal{G}$.

5.1. Voting by a committee

Consider $N = \{1,2,3\}$, $S_i = \{\phi, l, r\}$, $i \in N$. There are three players, each of which can vote for the status quo ϕ, left l, or right r. A change to the status quo occurs if and only if at least two players vote for it. Denote the possible outcomes by Φ for the status quo, L for a change to the left and R for a change to the right. The outcome of the vote is given by

$$o(s) = \begin{cases} L & \text{if} \quad |\{i \in N : s_i = l\}| \geq 2, \\ R & \text{if} \quad |\{i \in N : s_i = r\}| \geq 2, \\ \Phi & \text{otherwise.} \end{cases} \tag{24}$$

Each player will have one of two possible preferences, which will depend on outcomes. Independently across $i \in N$, player i has preference-type

$$t(i) = \begin{cases} l & \text{with probability } q, \\ r & \text{with probability } 1-q. \end{cases} \tag{25}$$

Payoffs are $\pi_i(s) = \rho_{t(i)}(o(s))$, where for constants $a, b > 0$,

$$\rho_l(L) = a, \qquad \rho_l(\Phi) = 0, \qquad \rho_l(R) = -b, \tag{26}$$

$$\rho_r(R) = a, \qquad \rho_r(\Phi) = 0, \qquad \rho_r(L) = -b. \tag{27}$$

Consequently, we have $2^{|N|} = 8$ possible $(\pi_i)_{i\in N}$ and eight games in $\mathcal{G}$.

To narrow down maxmin strategies, assume a lexicographic preference for voting for one's favorite outcome. That is, fixing the outcome, a player with preference-type l would rather vote l and a player with preference-type r would rather vote r.[9] Given this, individual maxmin strategies are $\underline{s}_i = t(i)$.

Now consider coalitional maxmin. First consider a coalition of size two, $T = \{i, j\}$. If $t(i) = t(j)$, then $MM(T) = \{(t(i), t(i))\}$ and coalition T ensures its favored outcome. If $t(i) \neq t(j)$, then $MM(T) = \{(\phi, \phi)\}$ and coalition T ensures that the least favored outcomes for each coalition member are avoided.

Now consider a coalition of size three, $T = \{i, j, k\}$. If $t(i) = t(j) = t(k)$, then $MM(T) = \{(t(i), t(i), t(i))\}$ and coalition T ensures its favored outcome. If $t(i) = t(j) \neq t(k)$, then $MM(T) = \{(t(i), t(i), t(k)), (t(i), \phi, t(k)), (\phi, t(j), t(k))\}$. One profile ensures the majority obtain their favored outcome, the other profiles ensure that the least favored outcomes for every player are avoided.[10,11]

Theorem 8. *In the model of voting by a committee, under binomial matching, $x = 1$ is an ESS, $x = 0$ is an ESS if and only if $a > b$, and there are no interior ESS.*

By playing coalitional maxmin, a player can avoid negative outcomes. However, collaborating with a player with opposed preferences can sometimes lead to an outcome that is worse than that which would have happened under individual maxmin. When there is a large share of α-types in the population, the former effect dominates. Consequently, there is an ESS at $x = 1$. When there are few α-types in the population, it turns out that the important coalitions are those formed of two α-types with opposed preferences.[12] By participating in such a coalition, a player ensures outcome Φ, whereas if the coalition had not formed, one α-type would have obtained his most favored outcome and the other α-type would have obtained his least favored outcome. Therefore, an ESS at $x = 0$ only exists if the loss from outcome Φ relative to the preferred outcome $(0 - a = -a)$ outweighs the gain from Φ relative to the least desired outcome $(0 - (-b) = b)$.

[9] This preference can be incorporated directly into payoffs via a small bonus payoff $\epsilon > 0$ when voting for one's favored outcome, but this does not change results.

[10] Note that profiles such as (ϕ, ϕ, ϕ) that give outcome Φ are not in $MM(T)$, as they are lexicographically dominated by profiles such as $(t(i), \phi, t(k))$ that also lead to Φ but allow some players to vote for their favored outcomes.

[11] Given that players in a coalition are engaged in some degree of knowledge pooling regarding their preferences, we may wish to think of $(t(i), t(i), t(k))$ as more likely than the other alternatives, as once players i and j recognize that they are in the majority, it is in their interests to enforce their will. However, such an assumption is unnecessary for Theorem 8.

[12] When there are few α-types in the population, coalitions of size three rarely occur. Coalitions formed of two players with identical preferences obtain the same outcome that would have been obtained under individual maxmin.

Example 2. To illustrate the importance of the matching protocol, consider highly assortative matching. For small ϵ, let $Pr_x[Z = 2 \mid \alpha]$ and $Pr_x[Z = 0 \mid \nu]$ be greater than $1 - \epsilon$ for all $x \in (0, 1)$, so that players are usually in homogeneous groups of the same type. If the expected payoff of α-types in such groups differs from the expected payoff of ν-types in such groups, then, for small enough ϵ, the comparison of these two payoffs suffices to determine ESS. Consider the case in which there is nonzero probability that three α-types with differing preferences induce outcome Φ. When this is the case, the comparison of payoffs between α-types and ν-types depends on comparing the expected payoff loss of inducing Φ when one's preferences are in the majority ($\frac{2}{3}a$) to the expected payoff gain of inducing Φ when one's preferences are in the minority ($\frac{1}{3}b$). If $b > 2a$, then there is a unique ESS $x = 1$. If $b < 2a$, then there is a unique ESS $x = 0$. Note that this latter case contrasts with the result under binomial matching [Theorem 8]. Assortativity in matching has led to $x = 1$ no longer being an ESS. In contrast with the evolution of cooperation literature, in which positive assortativity promotes cooperation, the evolution of coalitional behavior can be hindered by positive assortativity in matching.

6. Conclusion

The evolution of collaboration literature is still in its infancy. There are a wealth of applications and trajectories still to be studied. The current paper explores the evolution of collaboration against a background of maxmin choice. Naturally, this exploration leads to further questions and subtleties. In Appendix B we discuss[13]

1. Sufficient conditions for an ESS at $x = 0$.
2. The possibility of cheating / faulty collaborators, who fail to carry out their part in coalitional maxmin, whether by intent or by error.
3. Anti-free riding collaborators, who only participate in coalitional maxmin that includes every player.
4. Robustness to smaller sets of collaborators that might deviate from a larger set of collaborators.
5. Individualistic best responders, who correctly predict play and best respond to this prediction.

In this study, we have considered the evolution of collaboration in situations in which decision making is driven by maxmin considerations. We showed in Theorem 1 that when coalitional maxmin behavior leads to realized payoff gains, then collaboration proliferates when rare — any ESS has a nonzero share of collaborative types. Moving beyond this broad general result, we explored several economic applications, showing how stable shares of collaborative types in the population vary with payoffs and can exhibit non-monotonicity in payoff parameters due to thresholds at which coalitional behavior becomes viable for coalitions of given sizes.

Many questions remain. Relevant to the current paper's focus on collaboration and maxmin — a conservative strategy selection procedure, one might wonder whether conservatism helps or hinders collaboration. For example, when it comes to collaboration, how does maxmin compare to maxmax? If a manager in an organization wishes to encourage collaboration, what kind of strategy selection procedures (optimistic, pessimistic, conservative, speculative) should be encouraged? This and other questions are left for future research.

Declaration of competing interest

Jonathan Newton has received research grants from JSPS KAKENHI Grants-in-Aid for Scientific Research No. 21H00695. Miharu Naono has received no funding related to the research. We confirm that there have been no involvements that might raise the question of bias in the work reported. All financial support for this research is identified in the title page of the manuscript.

Appendix A. Proofs

A.1. Proof of preliminary lemma

In this subsection, we prove Lemma 1. This lemma states that if there exists a coalitional maxmin for a coalition T, then there also exists a coalitional maxmin for larger coalitions $T' \supset T$. This gives some monotonicity even in complicated situations, such as when the game is not symmetric or $G_{N_\alpha,\Gamma}(\cdot)$ puts positive probability on multiple strategy profiles. This is helpful for our proof of Theorem 1.

For the proof, we construct a dominance relation based on coalitional maxmin. This relation gives a partial ordering. We then show the existence of a maximal element under this order on the set of possible strategy profiles for the larger coalition (T') and that such an element is in $MM(T')$.

Proof of Lemma 1.

The first step is to define a partial ordering for any given $T \neq \emptyset$.

Partial ordering. Note that T is strictly α-effective for $\pi_T \in \Pi(s_T)$ if and only if there exists $s'_T \in S'_T$ such that, for all $\pi'_T \in \Pi_T(s'_T)$, $\pi'_T \geq \pi_T$. When this holds, we say that s'_T dominates s_T. It follows from this definition that if s'_T dominates s_T, then s_T does not

[13] We thank three referees for their comments and suggestions in this regard. In addition, we emphasize that the treatment of these extensions is far from comprehensive. Indeed, several of them would easily lend themselves to fuller treatment as research papers in their own right.

dominate s'_T. Furthermore, if s''_T dominates s'_T and s'_T dominates s_T, then s''_T dominates s_T. Therefore, this relation is a strict partial ordering on S_T.

Compactness and boundedness. Compactness of S and S_T follow from compactness of S_i, $i \in N$. Compactness of S together with continuity of π implies that payoffs $\pi_T(s)$, $s \in S$ are bounded above.

With these preliminaries complete, we proceed to show that any chain in S_T under this partial ordering is bounded above.

Chain $Z \subseteq S_T$ with ordering represented in one dimension. For $s_T \in S_T$, define

$$u_T(s_T) := \min\left\{ \sum_{i \in T} \pi_i(s_T, s_{-T}) : s_{-T} \in S_{-T} \right\}, \qquad \left[\begin{array}{c}\text{well-defined by}\\ \text{compactness of } S_{-T}\\ \text{and continuity of } \pi.\end{array}\right]. \tag{28}$$

If s''_T dominates s'_T, then there exists $\pi'_T \in \Pi_T(s'_T)$ such that for all $\pi''_T \in \Pi_T(s''_T)$, $\pi''_T \geq \pi'_T$. Therefore, $u_T(s''_T) > u_T(s'_T)$.

Let $Z \subseteq S_T$ be a chain (a totally ordered subset) under the dominance relation. As Z is totally ordered, we have that for any $s'_T, s''_T \in Z$, $u_T(s''_T) > u_T(s'_T)$ if and only if s''_T dominates s'_T. Furthermore, compactness of S_T implies that $\max_{s_T \in S_T} u_T(s_T)$ exists, thus $\sup_{s_T \in Z} u_T(s_T) =: \bar{u}$ is finite.

Chain $Z \subseteq S_T$ is bounded above. We will show that there exists $s_T \in S_T$ that is an upper bound on Z in that it dominates all elements of Z except for itself (if it is in Z). If Z has a maximal element, then this element is an upper bound. The next two steps show that an upper bound also exists when Z has no maximal element.

Step 1: Z is bounded in the limit by a countable sequence.

Choose a countable sequence of elements of Z, $\{s^t_T\}_{t \in \mathbb{Z}_+}$, such that $u_T(s^t_T)$ is increasing and $u_T(s^t_T) \geq \bar{u} - \frac{1}{2^t}$. For all $t \geq 1$, s^t_T dominates s^{t-1}_T, so there exists $\pi^{t-1}_T \in \Pi_T(s^{t-1}_T)$ such that for all $\pi_T \in \Pi_T(s^t_T)$, $\pi_T \geq \pi^{t-1}_T$. Note that $(\pi^t_T)_{t \in \mathbb{Z}_+}$ is an increasing sequence. As it is bounded above, it has a limit π^*_T. As the sequence is infinite, it must be that $\pi^t_T \leq \pi^*_T$ for all $t \in \mathbb{Z}_+$. Note that, for all $s_T \in Z$, $u_T(s_T) < \bar{u}$, otherwise s_T would be a maximal element of Z. Therefore, there exists τ such that, for all $t \geq \tau$, $u_T(s^t_T) > u_T(s_T)$ and, therefore, s^t_T dominates s_T.

Step 2: Countable sequence converges to an upper bound.

The Bolzano-Weierstrass theorem implies that $\{s^t_T\}_{t \in \mathbb{Z}_+}$ has a convergent subsequence. Denote the limit of some such subsequence by $s^*_T \in S_T$. By continuity of π, payoffs on the subsequence converge in the sense that, for given s_{-T}, $\pi_T(s^t_T, s_{-T}) \to \pi_T(s^*_T, s_{-T})$. For any s_{-T}, if $\pi_T(s^*_T, s_{-T}) \not\geqq \pi^*_T$, then for some t, we have $\pi_T(s^t_T, s_{-T}) \in \Pi_T(s^t_T)$, $\pi_T(s^t_T, s_{-T}) \not\geq \pi^{t-1}_T$, a contradiction. Therefore, $\pi_T \in \Pi_T(s^*_T)$ implies $\pi_T \geqq \pi^*_T$. From the previous step, we know that for all $t \in \mathbb{Z}_+$, $\pi^t_T \leq \pi^*_T$. Combining, we see that s^*_T dominates s^t_T for all $t \in \mathbb{Z}_+$.

We have shown that, for all $T \neq \emptyset$, any chain in S_T under our partial ordering has an upper bound. Next, we consider a given T (as in the lemma statement), $MM(T) \neq \emptyset$ and $T' \supset T$, then show that the set of strategy profiles $A_{T'}$ for T' that assure improvements upon individual maxmin are nonempty.

Payoff improvement on individual maxmin. By Assumption 1, there exists $T \subseteq N$, $|T| \geq 2$, $MM(T) \neq \emptyset$. Let $s_T \in MM(T)$. Let $T \subset T'$ and $s'_{T'}$ be such that $s'_T = s_T$, $s'_i \in \underline{S}_i$ for $i \in T' \setminus T$. Let $\pi'_{T'} \in \Pi_{T'}(s'_{T'})$. As $s'_i \in \underline{S}$ for $i \in T' \setminus T$, we have $\pi'_i \geq \underline{\pi}_i$. As $s'_T = s_T$, we have $\pi'_T \in \Pi_T(s_T)$, which together with $s_T \in MM(T)$, implies that $\pi'_T \geq \underline{\pi}_T$. Therefore, we have $\pi'_{T'} \geq \underline{\pi}_{T'}$. Thus, $A_{T'} := \{s_{T'} \in S_{T'} : \forall \pi_{T'} \in \Pi_{T'}(s_{T'}), \pi_{T'} \geq \underline{\pi}_{T'}\} \neq \emptyset$.

The final step is to apply Zorn's lemma to chains on $A_{T'}$ to obtain a profile that is a coalitional maxmin for T'.

Zorn's Lemma. As every chain in the partial ordering on $S_{T'}$ has an upper bound, every chain on $A_{T'}$ has an upper bound. Therefore, by Zorn's Lemma, $A_{T'}$ contains a maximal element under the partial ordering. That is, there exists at least one profile $s^{**}_{T'} \in A_{T'}$ that is not dominated. That is, T' is not strictly α-effective for any $\pi_{T'} \in \Pi_T(s^{**}_{T'})$. By definition of $MM(T')$, we have that $s^{**}_{T'} \in MM(T')$. □

A.2. Preliminaries for theorem proofs

In this subsection we give analytically useful expressions (π^k_α and π^k_ν) for the expected payoffs of α and ν-type players, conditional on how many α-type players are in a game. These expectations then combine with matching probabilities to give fitness functions $f_\alpha(x)$ and $f_\nu(x)$.

A.2.1. Expected payoffs

Denote the expected payoff of $i \in N$ at individual maxmin profiles, and the average over all players, by

$$\underline{\pi}^{avg}_i := \int_S \pi_i(s)\, d\, G_\Gamma(\cdot), \qquad \underline{\pi}^{avg} = \frac{1}{n} \sum_{i \in N} \underline{\pi}^{avg}_i. \tag{29}$$

Let $\pi_i(T)$ denote the expected payoff of $i \in N$ given that the set of α-types is $T \subseteq N$,

$$\pi_i(T) := \begin{cases} \int_S \pi_i(s)\, d\, G_{T,\Gamma}(\cdot) & \text{if } MM(T) \neq \emptyset \\ \underline{\pi}^{avg}_i & \text{if } MM(T) = \emptyset \end{cases} \tag{30}$$

Let $\mathcal{T}^k_i = \{T \subseteq N : |T| = k, i \in T\}$. Denote the average payoff of $i \in N$ conditional on being one of k α-types by $\pi^k_{\alpha,i}$, and the average over all players by π^k_α,

$$\pi^k_{\alpha,i} := \frac{1}{|\mathcal{T}_i^k|}\sum_{T\in\mathcal{T}_i^k}\pi_i(T), \qquad \pi^k_\alpha = \frac{1}{n}\sum_{i\in N}\pi^k_{\alpha,i}. \tag{31}$$

An α-type player will either play coalitional maxmin (when $MM(T)\neq\emptyset$) or play individual maxmin (when $MM(T)=\emptyset$). Writing $\mathcal{T}_i^{k+}=\{T\subseteq\mathcal{T}_i^k:\ MM(T)\neq\emptyset\}$, $\mathcal{T}_i^{k-}=\{T\subseteq\mathcal{T}_i^k:\ MM(T)=\emptyset\}$, substitute from (30) to obtain

$$\pi^k_{\alpha,i} = \frac{1}{|\mathcal{T}_i^k|}\left(\sum_{T\in\mathcal{T}_i^{k+}}\int_S \pi_i(s)\,d\,G_{T,\Gamma}(\cdot) + \sum_{T\in\mathcal{T}_i^{k-}}\underline{\pi}_i^{avg}\right). \tag{32}$$

The strategies of α-type players may also affect the payoffs of ν-type players. For ν-types, define $\pi^k_{\nu,i}$ π^k_ν similarly to $\pi^k_{\alpha,i}$ π^k_α by replacing $\mathcal{T}_i^k$ with $\overline{\mathcal{T}}_i^k=\{T\subseteq N:\ |T|=k,\ i\notin T\}$.

$$\pi^k_{\nu,i} := \frac{1}{|\overline{\mathcal{T}}_i^k|}\sum_{T\in\overline{\mathcal{T}}_i^k}\pi_i(T), \qquad \pi^k_\nu = \frac{1}{n}\sum_{i\in N}\pi^k_{\nu,i}. \tag{33}$$

Writing $\overline{\mathcal{T}}_i^{k+}=\{T\subseteq\overline{\mathcal{T}}_i^k:\ MM(T)\neq\emptyset\}$, $\overline{\mathcal{T}}_i^{k-}=\{T\subseteq\overline{\mathcal{T}}_i^k:\ MM(T)=\emptyset\}$,

$$\pi^k_{\nu,i} = \frac{1}{|\overline{\mathcal{T}}_i^k|}\left(\sum_{T\in\overline{\mathcal{T}}_i^{k+}}\int_S \pi_i(s)\,d\,G_{T,\Gamma}(\cdot) + \sum_{T\in\overline{\mathcal{T}}_i^{k-}}\underline{\pi}_i^{avg}\right). \tag{34}$$

Denote the lowest size of a coalition with a coalitional maxmin by

$$m := \min\{|T|:\ \exists T\subseteq N,\ MM(T)\neq\emptyset\}, \tag{35}$$

and denote maximum and minimum expected payoffs of ν and α-types respectively, given that $k\geq m$, by

$$\pi^{max}_\nu := \max_{k\geq m}\pi^k_\nu, \qquad \pi^{min}_\alpha := \min_{k\geq m}\pi^k_\alpha.$$

A.2.2. Fitness

Given expected payoffs, fitness can be written as

$$f_\alpha(x)=\sum_{k=0}^{n-1}Pr_x[Z=k\,|\,\alpha]\,\pi^{k+1}_\alpha, \qquad f_\nu(x)=\sum_{k=0}^{n-1}Pr_x[Z=k\,|\,\nu]\,\pi^k_\nu.$$

As mentioned in the main body of the paper, it is assumed that $G_{N_\alpha,\Gamma}(\cdot)$ and $G_\Gamma(\cdot)$ are such that the above expectations exist. Note that $f_\alpha(x)$ and $f_\nu(x)$ depend continuously on the probabilities $Pr_x[Z=k\,|\,\alpha]$ and $Pr_x[Z=k\,|\,\nu]$, which in turn are continuous in x. Therefore, $f_\alpha(x)$ and $f_\nu(x)$ are continuous in x.

A.3. Proofs of results under (RG)

Lemma 3. $\pi^{min}_\alpha - \underline{\pi}^{avg} > 0$.

Proof of Lemma 3.

Step 1: Weak inequality for all $i\in N$

Recall that $G_{T,\Gamma}$ puts positive weight only on s such that $s_T\in MM(T)$, $s_{-T}\in\underline{S}_{-T}$.

Therefore, $s\in\operatorname{supp}G_{T,\Gamma}$ together with (RG) implies $\pi_T(s)\geq\pi_T(\underline{s})$ for all $\underline{s}\in\underline{S}$.

Therefore, $\pi_i(s)\geq\pi_i(\underline{s})$ for all $i\in T$, $\underline{s}\in\underline{S}$.

It follows that $\pi_i(s)\geq\underline{\pi}_i^{avg}$ for all $i\in T$.

Therefore, (32) implies that $\pi^k_{\alpha,i}\geq\underline{\pi}_i^{avg}$ for all $i\in N$ and $k=1\dots,n$.

Step 2: For given k, strict inequality for some $i\in N$

By (35) and Lemma 1, for $k\geq m$, there exists T, $|T|=k$, such that $MM(T)\neq\emptyset$. For given k, choose such a T.

Step 1 showed that $\pi_i(s)\geq\underline{\pi}_i^{avg}$ for all $s\in\operatorname{supp}G_{T,\Gamma}$.

Together with (32), this implies that, if $\pi^k_{\alpha,i}=\underline{\pi}_i^{avg}$ for some $i\in T$, then the set of s such that $\pi_i(s)\neq\underline{\pi}_i^{avg}$ has zero measure under $G_{T,\Gamma}$.

So, with probability one, $G_{T,\Gamma}$ selects s such that $\pi_i(s)=\underline{\pi}_i^{avg}$ for all $i\in T$.

As $\underline{\pi}_i^{avg}$ is a weighted average over $\underline{S}$, this implies $\pi_T(s)\not\geq\pi_T(\underline{s})$ for all $\underline{s}\in\underline{S}$, contradicting (RG).

Step 3: Combining From Step 1, for all $i\in N$ and $k=1\dots,n$, we have $\pi^k_{\alpha,i}\geq\underline{\pi}_i^{avg}$. From Step 2, for some $i\in N$ and $k=m\dots,n$, we have $\pi^k_{\alpha,i}>\underline{\pi}_i^{avg}$. Averaging over i, for $k=m\dots,n$, we have $\pi^k_\alpha>\underline{\pi}^{avg}$. The definition of π^{min}_α completes the proof. □

Proof of Theorem 1.

The average fitness of an ν type is bounded above by

$$f_\nu(x) \le \underbrace{Pr_x[Z < m \mid \nu]}_{\substack{\text{Prob. too few}\\ \alpha \text{ types for}\\ \text{coalitional maxmin}}} \underline{\pi}^{avg} + \underbrace{Pr_x[Z \ge m \mid \nu]}_{\substack{\text{Prob. enough}\\ \alpha \text{ types for}\\ \text{coalitional maxmin}}} \pi_\nu^{max}.$$

The average fitness of an α type is bounded below by

$$f_\alpha(x) \ge \underbrace{Pr_x[Z < m-1 \mid \alpha]}_{\substack{\text{Prob. too few}\\ \alpha \text{ types for}\\ \text{coalitional maxmin}}} \underline{\pi}^{avg} + \underbrace{Pr_x[Z \ge m-1 \mid \alpha]}_{\substack{\text{Prob. enough}\\ \alpha \text{ types for}\\ \text{coalitional maxmin}}} \pi_\alpha^{min}.$$

Subtracting,

$$f_\alpha(x) - f_\nu(x) = (f_\alpha(x) - \underline{\pi}^{avg}) - (f_\nu(x) - \underline{\pi}^{avg}) \tag{36}$$

$$\ge Pr_x[Z \ge m-1 \mid \alpha](\pi_\alpha^{min} - \underline{\pi}^{avg}) - Pr_x[Z \ge m \mid \nu](\pi_\nu^{max} - \underline{\pi}^{avg}).$$

The balance condition for matchings (5) implies that for $k \ge 1$,

$$\frac{Pr_x[Z \ge m-1 \mid \alpha]}{Pr_x[Z \ge m \mid \nu]} \to \infty \quad \text{as} \quad x \to 0. \tag{37}$$

Therefore, for small enough x,

$$Pr_x[Z \ge m-1 \mid \alpha] > Pr_x[Z \ge m \mid \nu] \left(\frac{\pi_\nu^{max} - \underline{\pi}^{avg}}{\pi_\alpha^{min} - \underline{\pi}^{avg}} \right), \tag{38}$$

Lemma 3 implies that $\pi_\alpha^{min} - \underline{\pi}^{avg} > 0$. Together with (38), this implies that the RHS of (36) is greater than zero for small enough x. That is, $x = 0$ cannot be an ESS, so any ESS must have $x > 0$. □

Proof of Corollary 1.

Step 1. $MM(N_\alpha) = \varnothing$ for $N_\alpha \ne N$. Therefore, for all $i \in N$, $\overline{\mathcal{T}}_i^{k+} = \varnothing$ for $k = 0, \dots, n-1$, so (34) implies that $\pi_{\nu,i}^k = \underline{\pi}_i^{avg}$. Therefore, $\pi_\nu^k = \underline{\pi}^{avg}$ for $k = 1, \dots, n-1$. Consequently, $f_\nu(x) = \underline{\pi}^{avg}$ for all x.

Step 2. Similarly, for all $i \in N$, $\mathcal{T}_i^{k+} = \varnothing$ for $k = 1, \dots, n-1$, so (32) implies that $\pi_{\alpha,i}^k = \underline{\pi}_i^{avg}$. Therefore, $\pi_\alpha^k = \underline{\pi}^{avg}$ for $k = 1, \dots, n-1$.

Step 3. For $k = n$, $\mathcal{T}_i^{k+} \ne \varnothing$. Lemma 3 implies that $\pi_\alpha^{min} = \min_{k \ge m} \pi_\alpha^k > \underline{\pi}^{avg}$. This, together with Step 2, implies that $\pi_\alpha^n > \underline{\pi}^{avg}$. Thus, for all x,

$$f_\alpha(x) - f_\nu(x) = (f_\alpha(x) - \underline{\pi}^{avg}) - \underbrace{(f_\nu(x) - \underline{\pi}^{avg})}_{=0 \text{ by Step 1}} \tag{39}$$

$$= \sum_{k=0}^{n-1} Pr_x[Z = k \mid \alpha]\, \pi_\alpha^{k+1} - \underline{\pi}^{avg}$$

$$= Pr_x[Z = n-1 \mid \alpha] \underbrace{(\pi_\alpha^n - \underline{\pi}^{avg})}_{>0 \text{ by Step 3}} + \sum_{k=0}^{n-2} Pr_x[Z = k \mid \alpha] \underbrace{(\pi_\alpha^{k+1} - \underline{\pi}^{avg})}_{=0 \text{ by Step 2}} > 0. \quad \square$$

A.4. Proofs of results on social dilemmas

Proof of Lemma 2.

Consider $\underline{s}, \underline{s}' \in \underline{S}$ and $s_T \in MM(T)$. By definition of $MM(T)$, $s_T \in MM(T)$ implies that $\forall \pi_T \in \Pi(s_T), \pi_T \ge \underline{\pi}_T$. In particular, $\pi_T(s_T, \underline{s}_{-T}) \in \Pi(s_T)$, therefore $\pi_T(s_T, \underline{s}_{-T}) \ge \underline{\pi}_T$. Further, (SD) implies $\underline{\pi}_T = \pi_T(\underline{s}')$, so $\pi_T(s_T, \underline{s}_{-T}) \ge \pi_T(\underline{s}')$. □

Proof of Theorem 2.

Note that there is only one profile in $\underline{S}$. At this profile $s_i = 0$ for all i. It follows that $\underline{\pi}^{avg} = 0$. Further, from the definition of m in (35)

$$m = \left\lfloor \frac{c}{b} + 1 \right\rfloor \ge 2. \tag{40}$$

If $m < n$, then

$$f_v(x) = \sum_{k=m}^{n-1} Pr_x[Z = k]\, kb \tag{41}$$

$$f_\alpha(x) = \sum_{k=m-1}^{n-1} Pr_x[Z = k]\, ((k+1)b - c)$$

and

$$f_\alpha(x) - f_v(x) = Pr_x[Z = m-1](mb - c) + \sum_{k=m}^{n-1} Pr_x[Z = k](b - c) \tag{42}$$

Dividing by $Pr_x[Z = m-1] > 0$, we obtain

$$(mb - c) + (b - c) \sum_{k=m}^{n-1} \frac{Pr_x[Z = k]}{Pr_x[Z = m-1]} \tag{43}$$

$$= \underbrace{(mb - c)}_{>0} + \underbrace{(b - c)}_{<0} \underbrace{\sum_{k=m}^{n-1} \underbrace{\left(\prod_{l=m}^{k} \frac{n-l+1}{l} \right)}_{>0} \underbrace{\left(\frac{x}{1-x} \right)^{k-m+1}}_{>0}}_{=:A<0} \tag{44}$$

Proof of (i).

m<n. (44) is strictly decreasing in x. As $x \to 0$, $A \to 0$, so (44) > 0. As $x \to 1$, $A \to -\infty$, so (44) < 0. This implies a unique ESS x^* such that $0 < x^* < 1$.

m = n. Applying Corollary 1, we have a unique ESS $x^* = 1$.

Proof of (ii).

Consider the replicator equation

$$\dot{x} = x\left(f_\alpha(x) - (xf_\alpha(x) + (1-x)f_v(x))\right) = x(1-x)(f_\alpha(x) - f_v(x)). \tag{45}$$

It follows from the proof of (i) that, for $x \in (0,1)$, $x < x^*$ implies $\dot{x} > 0$ and $x > x^*$ implies $\dot{x} < 0$.

Proof of (iii) — *varying n*.

m<n. Each term under the $\sum$ in (44) strictly increases in n. Furthermore, increasing n adds additional positive terms to the summation. Thus the sum strictly increases in n. This implies (44) strictly decreases in n. Thus, higher n implies that (44) is equal to zero at a strictly lower x^*.

m = n. Note that, apart from being bounded above by n, m does not depend on n. Therefore, from a starting point of $m = n$, increasing n leads to $m < n$, so (i) implies that x^* strictly decreases.

Proof of (iii) — *varying c*.

Non-monotonicity in c is implied by (v), proved below.

Proof of (iv).

n = 2. (40) implies $m \geq 2$, so we have $m = n$. Corollary 1 implies that $x^* = 1$.

n>2. As $c \downarrow b$, (40) implies $m \to 2$. Therefore, for given $x \in (0,1)$, (44) $\to b > 0$. As (44) is strictly decreasing in x, this implies $x^* \to 1$.

Proof of (v).

For fixed $m < n$, (44) is strictly decreasing in c, therefore x^* is strictly decreasing in c. At $c = pb$, $2 \leq p \leq n-1$, m increases by 1.

As $c \uparrow pb$, (40) implies $m \to p$. Therefore, for given $x \in (0,1)$,

$$(44) \to b \underbrace{(1-p)}_{<0} \underbrace{\sum_{k=p}^{n-1} \left(\prod_{l=k}^{p} \frac{n-l+1}{l} \right) \left(\frac{x}{1-x} \right)^{k-p+1}}_{>0} < 0. \tag{46}$$

As (44) is strictly decreasing in x, this implies $x^* \to 0$ as $c \uparrow pb$.

At $c = pb$, (40) implies $m = p+1$. If $p+1 = n$, then Corollary 1 implies that $x^* = 1$. If $p+1 < n$, then (i) implies that $x^* \in (0,1)$. That is, $x^* > 0$.

Proof of (vi).

If $c \geq (n-1)b$, then (40) implies $m = n$, so Corollary 1 implies $x^* = 1$. □

Proof of Theorem 3.

Note that there is only one profile in $\underline{S}$. At this profile $s_i = 0$ for all i. It follows that $\underline{\pi}^{avg} = 0$.

If $m < n$, then

$$f_\nu(x) = \sum_{k=m}^{n-1} Pr_x[Z = k]\, b \tag{47}$$

$$f_\alpha(x) = \sum_{k=m-1}^{n-1} Pr_x[Z = k] \left(b - c\frac{m}{k+1}\right),$$

with the $m/k+1$ multiplier arising because only m of the players in a coalition contribute as part of coalitional maxmin. It follows that

$$f_\alpha(x) - f_\nu(x) = Pr_x[Z = m-1]\, b - \sum_{k=m-1}^{n-1} Pr_x[Z = k]\, c\frac{m}{k+1} \tag{48}$$

Dividing by $c\, Pr_x[Z = m-1] > 0$, we obtain

$$\frac{b}{c} - \sum_{k=m-1}^{n-1} \frac{Pr_x[Z = k]}{Pr_x[Z = m-1]} \frac{m}{k+1} \tag{49}$$

$$= \underbrace{\frac{b}{c}}_{>1} - 1 - \underbrace{\sum_{k=m}^{n-1} \left(\prod_{l=m}^{k} \frac{n-l+1}{l} \right) \left(\frac{x}{1-x}\right)^{k-m+1}}_{=:B>0} \tag{50}$$

Proof of (i).

m*<*n. (50) is strictly decreasing in x. As $x \to 0$, $B \to 0$, so (50) > 0. As $x \to 1$, $B \to \infty$, so (50) < 0. This implies a unique ESS x^* such that $0 < x^* < 1$.

m* = *n. Applying Corollary 1, we have a unique ESS $x^* = 1$.

Proof of (ii).

Consider the replicator equation

$$\dot{x} = x\left(f_\alpha(x) - (xf_\alpha(x) + (1-x)f_\nu(x))\right) = x(1-x)(f_\alpha(x) - f_\nu(x)). \tag{51}$$

It follows from the proof of (i) that, for $x \in (0,1)$, $x < x^*$ implies $\dot{x} > 0$ and $x > x^*$ implies $\dot{x} < 0$.

Proof of (iii) — *varying n.*

m*<*n. Each term under the $\sum$ in (50) strictly increases in n. Furthermore, increasing n adds additional positive terms to the summation. Thus the sum strictly increases in n. This implies (50) strictly decreases in n. Thus, higher n implies that (50) is equal to zero at a strictly lower x^*.

m* = *n. increasing n leads to $m < n$, so (i) implies that x^* strictly decreases.

Proof of (iii) — *varying* b/c.

m*<*n. (50) strictly increases in b/c. Thus, higher b/c implies that (50) is equal to zero at a strictly higher x^*.

m* = *n. Corollary 1 implies that for any $b/c > 1$, we have a unique ESS $x^* = 1$.

Proof of (iv).

For given $x \in (0,1)$, as $b/c \to 1$, (50) $\to -B < 0$. As (50) is strictly decreasing in x, this implies $x^* \to 0$ as $b/c \to 1$.

For given $x \in (0,1)$, as $b/c \to \infty$, (50) $\to \infty$. As (50) is strictly decreasing in x, this implies $x^* \to 1$ as $b/c \to \infty$. □

Proof of Theorem 4.

(SD) and Lemma 2 imply (RG) holds. Furthermore, (13) implies $MM(T) \neq \emptyset \Leftrightarrow T = N$. Therefore, (C) and Corollary 1 imply a unique ESS $x^* = 1$. □

A.5. Proofs of results on price competition

Proof of Theorem 5.

From our discussion in the main text, we know

- $\underline{s} \in \underline{S}$ implies $\underline{s}_i = \frac{D}{2(a-b)}$ for all $i \in N$.
- $s_T \in MM(T)$ implies $s_i = \frac{D}{2(a-b|T|)}$ for all $i \in T$.

For $|T| < 2$, $MM(T)$ is empty, so

$$\pi_\nu^0 = \pi_\nu^0 = \pi_\alpha^1 = \pi_i(\underline{s}), \qquad i \in N. \tag{52}$$

For $|T| = k \geq 2$,

$$\pi_\alpha^k = \pi_i(s_T, \underline{s}_{-T}), \quad i \in T, \tag{53}$$

$$\pi_{\nu}^{k} = \pi_i(s_T, \underline{s}_{-T}), \quad i \notin T. \tag{54}$$

Substituting strategies $\underline{s}_i$, s_T into the demand and then payoff function, noting that the zero lower bound on $D_i(\cdot)$ is never attained,

$$\pi_{\alpha}^{k} = \left(D - a\frac{D}{2(a-k\,b)} + b\left((n-k)\frac{D}{2(a-b)} + k\frac{D}{2(a-k\,b)}\right)\right)\frac{D}{2(a-k\,b)}, \tag{55}$$

$$\pi_{\nu}^{k} = \left(D - a\frac{D}{2(a-b)} + b\left((n-k)\frac{D}{2(a-b)} + k\frac{D}{2(a-k\,b)}\right)\right)\frac{D}{2(a-b)}, \tag{56}$$

Note that $\pi_{\alpha}^{1} - \pi_{\nu}^{0} = 0$. For $k = 1, \ldots, n-1$, rearranging

$$\pi_{\alpha}^{k+1} - \pi_{\nu}^{k} = \frac{b^2 D^2 k\,((a-bk)(n-k) + b(k-1))}{4(a-b)^2(a-bk)(a-b(1+k))} > 0 \tag{57}$$

Proof of (i).

For $s_T \in MM(T)$, $|T| = k+1$, $k = 1, \ldots, n-1$,

$$\pi_i(s_T, \underline{s}_{-T}) \underbrace{=}_{\text{by (53)}} \pi_{\alpha}^{k+1} \underbrace{>}_{\text{by (57)}} \pi_{\nu}^{k} \underbrace{\geq}_{\text{by (56)}} \pi_{\nu}^{0} \underbrace{=}_{\text{by (52)}} \pi_i(\underline{s})$$

which shows (RG) for the set $MM(T)$ restricted to the symmetric coalitional minmax strategies that we consider.

Therefore, by Theorem 1, we have $x > 0$ at any ESS.

Proof of (ii).

Under binomial matching, for $x \in (0,1)$, we have

$$f_{\alpha}(x) - f_{\nu}(x) = \sum_{k=0}^{n-1} \underbrace{Pr_x[Z=k]}_{>0 \text{ by } x\in(0,1)} \underbrace{(\pi_{\alpha}^{k+1} - \pi_{\nu}^{k})}_{\substack{=0 \text{ for } k=0 \text{ by (52)} \\ >0 \text{ for } k\geq 1 \text{ by (57)}}} > 0, \tag{58}$$

implying a unique ESS $x^* = 1$. □

Proof of Theorem 6.

As discussed in the main text, $MM(T) = \varnothing$ for all $T \neq N$ and $MM(T) \neq \varnothing$ for $T = N$.

Furthermore, for all $i \in N$ individual maxmin payoffs are $\pi_i(\underline{s}) = 0$, whereas for $s \in MM(N)$, $T = N$, $\pi_i(s) > 0$. Therefore (RG) holds.

Therefore Corollary 1 implies there is a unique ESS $x^* = 1$. □

Proof of Theorem 7.

From our discussion in the main text, we know

- $\underline{s} \in \underline{S}$ implies $\underline{s}_i = \frac{D}{2(a-b)}$ for all $i \in N$.
- $s_T \in MM(T)$, $|T| < n$ implies $s_i = \frac{D}{2(a-b\,|T|)}$ for all $i \in T$.
- $s \in MM(N)$ implies $s_i = \frac{\theta\,D + (1-\theta)\,D'}{2\theta(a-b\,n)+2(1-\theta)}$ for all $i \in N$.

Substituting strategies $\underline{s}_i$, s_T into the payoff function, for $k < 2$,

$$\pi_{\alpha}^{k} = \pi_{\nu}^{k} = \theta\left(D - (a-nb)\frac{D}{2(a-b)}\right)\frac{D}{2(a-b)} + (1-\theta)\frac{1}{n}\left(D' - \frac{D}{2(a-b)}\right)\frac{D}{2(a-b)}. \tag{59}$$

For $2 \le k < n$,

$$\pi_\alpha^k = \theta\Bigg(D - a\frac{D}{2(a-k\,b)} + b\left((n-k)\frac{D}{2(a-b)} + k\frac{D}{2(a-k\,b)}\right)\Bigg)\frac{D}{2(a-k\,b)}, \tag{60}$$

$$\pi_\nu^k = \theta\Bigg(D - a\frac{D}{2(a-b)} + b\left((n-k)\frac{D}{2(a-b)} + k\frac{D}{2(a-k\,b)}\right)\Bigg)\frac{D}{2(a-b)} + (1-\theta)\frac{1}{n-k}\left(D' - \frac{D}{2(a-b)}\right)\frac{D}{2(a-b)}. \tag{61}$$

For $k = n$,

$$\pi_\alpha^k = \theta\left(D - (a-nb)\,\frac{\theta\,D + (1-\theta)\,D'}{2\theta(a-b\,n) + 2(1-\theta)}\right)\frac{\theta\,D + (1-\theta)\,D'}{2\theta(a-b\,n) + 2(1-\theta)} + (1-\theta)\frac{1}{n}\left(D' - \frac{\theta\,D + (1-\theta)\,D'}{2\theta(a-b\,n) + 2(1-\theta)}\right)\frac{\theta\,D + (1-\theta)\,D'}{2\theta(a-b\,n) + 2(1-\theta)}. \tag{62}$$

Proof of (i-a).
Under binomial matching, for $x \in (0,1)$, we have

$$f_\alpha(x) - f_\nu(x) = \sum_{k=0}^{n-1} Pr_x[Z=k] \underbrace{(\pi_\alpha^{k+1} - \pi_\nu^k)}_{=0 \text{ for } k=0 \text{ by (59)}} = \sum_{k=1}^{n-1} Pr_x[Z=k]\,(\pi_\alpha^{k+1} - \pi_\nu^k). \tag{63}$$

which has the same sign as

$$\sum_{k=1}^{n-1} \frac{Pr_x[Z=k]}{Pr_x[Z=1]}\,(\pi_\alpha^{k+1} - \pi_\nu^k) = (\pi_\alpha^2 - \pi_\nu^1) + \sum_{k=2}^{n-1} \underbrace{\frac{Pr_x[Z=k]}{Pr_x[Z=1]}}_{\to 0 \text{ as } x\to 0}\,(\pi_\alpha^{k+1} - \pi_\nu^k). \tag{64}$$

As $\theta \to 0$, (60) implies

$$\pi_\alpha^2 \to 0,$$

and (59) implies

$$\pi_\nu^1 \to \frac{1}{n}\left(D' - \frac{D}{2(a-b)}\right)\frac{D}{2(a-b)} > 0,$$

therefore for small enough θ, $\pi_\alpha^2 - \pi_\nu^1 < 0$.

It follows that as $x \to 0$, the expression in (64) becomes strictly negative, therefore (63) is also strictly negative and we have an ESS at $x = 0$.

Proof of (i-b). (63) has the same sign as

$$\sum_{k=1}^{n-1} \frac{Pr_x[Z=k]}{Pr_x[Z=n-1]}\,(\pi_\alpha^{k+1} - \pi_\nu^k) = \sum_{k=1}^{n-2} \underbrace{\frac{Pr_x[Z=k]}{Pr_x[Z=n-1]}}_{\to 0 \text{ as } x\to 1}\,(\pi_\alpha^{k+1} - \pi_\nu^k) + (\pi_\alpha^n - \pi_\nu^{n-1}). \tag{65}$$

As $\theta \to 0$, (62) implies

$$\pi_\alpha^n \to \frac{1}{n}\left(D' - \frac{D'}{2}\right)\frac{D'}{2} = \frac{D'^2}{4n},$$

and (61) implies

$$\pi_\nu^{n-1} \to \left(D' - \frac{D}{2(a-b)}\right)\frac{D}{2(a-b)},$$

therefore, if $D \ll D'$, then for small enough θ, $\pi_\alpha^n - \pi_\nu^{n-1} > 0$.

It follows that as $x \to 0$, the expression in (65) becomes strictly positive, therefore (63) is also strictly positive and we have an ESS at $x = 1$.

Proof of (i-c).

Again, considering (65), if $\frac{D}{2(a-b)} = \frac{D'}{2}$, then for small enough θ, $\pi_\alpha^n - \pi_\nu^{n-1} < 0$.

It follows that as $x \to 0$, the expression in (65) becomes strictly negative, therefore (63) is also strictly negative and there is no ESS at $x = 1$.

Proof of (ii).

By (59), $\pi_\alpha^1 = \pi_\nu^0$.

As $\theta \to 1$, the expressions for π_α^k, π_ν^k in (59) - (62) converge to the expressions in (55) - (56). Therefore, for $k = 1, \ldots, n-1$, (57) still holds and $\pi_\alpha^{k+1} - \pi_\nu^k > 0$.

Therefore, (58) also holds, implying a unique ESS $x^* = 1$. □

A.6. Proofs of results on voting on a committee

Proof of Theorem 8.

If all players play individual maxmin, then a player's payoff is a if he shares the preferences of at least one other player, and b if he does not. So for $k, k' < 2$,

$$\pi_\nu^k = \pi_\alpha^{k'} = a\underbrace{\left(1 - q(1-q)^2 - (1-q)q^2\right)}_{\substack{\text{prob given player has same} \\ t(\cdot) \text{ as at least one other player}}} a - \underbrace{\left(q(1-q)^2 + (1-q)q^2\right)}_{\substack{\text{prob given player has different} \\ t(\cdot) \text{ to both other players}}} b \tag{66}$$

$$= \left(1 - q + q^2\right) a - (q - q^2)\, b.$$

If $k = 2$, then a ν-type obtains a payoff of 0 when the two α-types have different preferences, and a or b when the two α-types have the same preferences.

$$\pi_\nu^2 = \underbrace{\left(q^3 + (1-q)^3\right)}_{\substack{\text{prob } \nu\text{-type has same} \\ t(\cdot) \text{ as both } \alpha\text{-types}}} a - \underbrace{\left(q(1-q)^2 + (1-q)q^2\right)}_{\substack{\text{prob } \nu\text{-type has different} \\ t(\cdot) \text{ to both } \alpha\text{-types}}} b \tag{67}$$

$$= \left(q^3 + (1-q)^3\right) a - (q - q^2)\, b$$

If $k = 2$, then an α-type obtains a payoff of a when the other α-type has the same preferences, and zero otherwise.

$$\pi_\alpha^2 = \underbrace{\left(q^2 + (1-q)^2\right)}_{\substack{\text{prob given } \alpha\text{-type has} \\ \text{same } t(\cdot) \text{ as other } \alpha\text{-type}}} a = \left(2q^2 - 2q + 1\right) a. \tag{68}$$

If $k = 3$, then an α-type obtains a payoff of a when the other two α-types have the same preferences. We saw in the main text that when the players have differing preferences, coalitional maxmin may lead to the majority's preferred outcome, or to the status quo. In general, given $G_{N,\Gamma}$, denote the probability that the majority opinion prevails in such cases by $\beta \in [0, 1]$. Then,

$$\pi_\alpha^3 = \underbrace{\left(q^3 + (1-q)^3\right)}_{\substack{\text{prob players all} \\ \text{have same } t(\cdot)}} a - \underbrace{\left(1 - q^3 - (1-q)^3\right)}_{\substack{\text{prob players do not} \\ \text{all have same } t(\cdot)}} \left(\beta \underbrace{\left(\frac{2}{3}a + \frac{1}{3}b\right)}_{\substack{\text{average payoff when} \\ \text{majority prevails}}}\right) \tag{69}$$

$$= \left(q^3 + (1-q)^3\right) a + (q - q^2)\, \beta(2a - b)$$

Under binomial matching,

$$f_\alpha(x) - f_\nu(x) = \sum_{k=0}^{2} Pr_x[Z = k]\left(\pi_\alpha^{k+1} - \pi_\nu^k\right) \tag{70}$$

$$= (1-x)^2 \underbrace{(\pi_\alpha^1 - \pi_\nu^0)}_{=0} + 2x(1-x) \underbrace{(\pi_\alpha^2 - \pi_\nu^1)}_{=(q-q^2)(b-a)} \tag{71}$$

$$+ x^2 \underbrace{(\pi_\alpha^3 - \pi_\nu^2)}_{=(q-q^2)(2\beta a+(1-\beta)b)} \tag{72}$$

As we wish to consider the sign of $f_\alpha(x) - f_\nu(x)$, we divide by $x(1-q)q > 0$ to get the continous function

$$F(x) := \frac{f_\alpha(x) - f_\nu(x)}{x(1-q)q} = 2(1-x)(b-a) + x(2\beta a + (1-\beta)b). \tag{73}$$

Proof of ESS at $\mathbf{x = 1}$.

As $F(1) = 2\beta a + (1-\beta)b > 0$, we have that $f_\alpha(x) - f_\nu(x)$ is strictly positive in some open interval bounded above by $x = 1$, therefore $x = 1$ is an ESS.

Proof of ESS at $\mathbf{x = 0 \Leftrightarrow a > b}$.

As $F(0) = 2(b-a)$, we have that $F(0) < 0$ if and only if $a > b$. Therefore, $f_\alpha(x) - f_\nu(x)$ is strictly negative in some open interval bounded below by $x = 0$, that is $x = 0$ is an ESS, if and only if $a > b$.

Proof of no ESS in $(\mathbf{0}, \mathbf{1})$.

$\frac{dF}{dx} = (\beta + 1)(2a - b)$, which is constant in x. Therefore, there is at most one $x' \in (0,1)$ for which $F(x') = 0$. However, $F(1) > 0$ implies that if such an x' exists, it must be that $\frac{dF}{dx}(x') > 0$, therefore x' is not an ESS. □

A.6.1. Details of Example 2

Write

$$J := \pi_\alpha^1 + \pi_\alpha^2 + \pi_\nu^0 + \pi_\nu^1. \tag{74}$$

By assumption, $Pr_x[Z = 2 \mid \alpha] > 1 - \epsilon$ and $Pr_x[Z = 0 \mid \nu] > 1 - \epsilon$ for all x. Consequently, the definitions of $f_\alpha(x)$ and $f_\nu(x)$ imply that

$$(1-\epsilon)(\pi_\alpha^3 - \pi_\nu^0) - \epsilon J \le f_\alpha(x) - f_\nu(x) \le (1-\epsilon)(\pi_\alpha^3 - \pi_\nu^0) + \epsilon J. \tag{75}$$

Calculating,

$$\pi_\alpha^3 - \pi_\nu^0 = (q - q^2)(1-\beta)(b - 2a), \tag{76}$$

which does not depend on x and, for $\beta \in [0,1)$, has the same sign as $b - 2a$.

(75) and (76) together imply that, for small enough ϵ, $f_\alpha(x) - f_\nu(x)$ has the same sign as $b - 2a$ for all x. Therefore, $b - 2a < 0$ implies a unique ESS at $x = 0$ and $b > 2a$ implies a unique ESS at $x = 1$.

Appendix B. Discussion

B.1. Conditions for ESS at $x = 0$

Theorem 1 gives sufficient conditions for the non-existence of an ESS at $x = 0$. It is natural to ask the reverse — under what conditions is there an ESS at $x = 0$? Here, we consider this with reference to Example 1 from earlier in the paper.

Example 1 (continued). Consider the game in Example 1. For $|T| = 2$, we have $MM(T) = \{(A, A)\}$, giving realized payoffs $(5, 5)$ for T. Further observe that $MM(N) = \{(A, A, A)\}$, with realized payoffs $(10, 10, 10)$.

The average realized payoff of ν-types is given by

$$\underbrace{Pr_x[Z = 0 \mid \nu]\, 6}_{\text{no collaboration}} + \underbrace{Pr_x[Z = 1 \mid \nu]\, 6}_{\text{no collaboration}} + \underbrace{Pr_x[Z = 2 \mid \nu]\, 2}_{\text{the other two players collaborate}}, \tag{77}$$

and the average realized payoff of α-types is given by

$$\underbrace{Pr_x[Z = 0 \mid \alpha]\, 6}_{\text{no collaboration}} + \underbrace{Pr_x[Z = 1 \mid \alpha]\, 5}_{\text{collaborates with one other player}} + \underbrace{Pr_x[Z = 2 \mid \alpha]\, 10}_{\text{collaborates with two other players}}. \tag{78}$$

Binomial matching. It can be checked that under binomial matching, for x close to 0, (77) is greater than (78), therefore there is an ESS at $x = 0$.

Positively assortative matching. We saw in Example 1 that the condition that (RG) imposes on $T \subseteq N$ is violated when $|T| = 2$. However, the condition in (RG) does hold when $T = N$. This suggests that there may exist a matching protocol under which α-types proliferate when rare. Consider a balanced matching protocol where we first match a share 0.9 of α-types into groups of three α-types,

then match the remainder of the population using binomial matching. This gives $Pr_x[Z = 2 \mid \alpha] > 0.9$, so by (78), the average payoff of α-types is greater than $0.9(10) = 9$, which is greater than the average payoff of ν-types for all $x \in (0, 1)$. Therefore, there is a unique ESS at $x = 1$. There is no ESS at $x = 0$.

The above suggests that if we wish to create games that have an ESS at $x = 0$ for any matching protocol, there should be no realized gains for any $T \subseteq N$. Indeed, if we consider a *realized losses* condition by replacing the $\geq$ in (RG) with $\leq$, then we can make an argument similar to the proof of Theorem 1. When α-types are a small share of the population, any given α-type will find himself in a group in which coalitional maxmin occurs much more frequently than any given ν-type finds himself in such a group. Thus an α-type will suffer realized payoff losses from collaboration far more often than ν-types are affected by collaboration, either positively or negatively. Consequently, there is an ESS at $x = 0$.

B.2. Cheating / faulty collaborators

Consider a population of ν-types and α-types at some ESS $x^* > 0$. We consider a novel type — the $\alpha^\dagger$-type, and ask whether x^* is vulnerable to invasion. Assume that $\alpha^\dagger$-types plan and coordinate coalitional maxmin together with α-types, but are unable to carry out their part of the plan, instead playing individual maxmin as if they were ν-types. Assume that $\alpha^\dagger$-types are otherwise indistinguishable from α-types. Let matching be binomial.

Consider the prisoner's dilemma (Section 3.1). We know from Rusch (2019) that it is difficult to sustain collaboration in the 2-player prisoner's dilemma with imperfect recognition of types. This extends to the n-player version. Specifically, when matched to $z \geq m - 1$ α-types, an

1. α-type will contribute and obtain a payoff of $(z + 1)b - c$.
2. $\alpha^\dagger$-type will not contribute and will obtain a payoff of zb.

Hence, the average payoff of $\alpha^\dagger$-types will exceed that of α-types (hence also ν-types, as x^* is an ESS). That is, x^* is vulnerable to invasion by $\alpha^\dagger$-types.[14]

For threshold public goods (Section 3.2), things are simpler. When matched to at least $m - 1$ α-types and selected to contribute as part of coalitional maxmin, an

1. α-type will contribute and obtain a payoff of $b - c > 0$.
2. $\alpha^\dagger$-type will not contribute and will obtain a payoff of 0.

Hence, the average payoff of $\alpha^\dagger$-types will be lower than that of α-types (hence also ν-types, as x^* is an ESS). That is, x^* is robust to invasion by $\alpha^\dagger$-types.

Overall, the key question is, for potentially cheating or faulty participants T' in some likely coalition T, conditional on $T \setminus T'$ continuing to do its part in a coalitional maxmin, whether the players in T' benefit from doing their part rather than playing individual maxmin. Of course, there are many subtleties inherent in this broad principle. For example, with asymmetric coalitional maxmin profiles, even if invasion of $\alpha^\dagger$-types is impossible, there might still be vulnerability to some sophisticated invader who acts like an α-type for some proposed coalitional maxmin profiles, but like an $\alpha^\dagger$-type for other proposed coalitional maxmin profiles.

B.3. Anti-free riding collaborators

Consider a population of ν-types and α-types at ESS $x^* \in (0, 1)$. We consider a novel type — the $\alpha^\varnothing$-type, and ask whether x^* is vulnerable to invasion. Assume that $\alpha^\varnothing$-types do not collaborate when some non-collaborator would be affected. That is, they are identical to ν-types except for the case in which every other player is a collaborative type. Assume binomial matching and that collaboration is possible for some strict subset of N, otherwise we are in the situation of Corollary 1.

Conditional on the types of other players, payoff differences for $i \in N$ when he is ν-type compared to $\alpha^\varnothing$-type only arise when $N \setminus \{i\}$ is composed of collaborative types. As we consider a small invasion of $\alpha^\varnothing$-types, it will usually be the case that these other collaborative players are α-types. Then $\alpha^\varnothing$-types can outperform ν-types (hence also α-types, as x^* is an ESS) if payoffs from being the nth member of a coalition exceed payoffs from free riding on an $n - 1$ member coalition.

It is also possible to compare payoffs for $i \in N$ when he is α versus $\alpha^\varnothing$-type. A difference arises when there is at least one ν-type in the player set. When that is the case, there may be situations in which an

1. α-type will collaborate together with other α-types.
2. $\alpha^\varnothing$-type will play individual maxmin.

[14] Of course, any resulting population can be invaded by smarter $\alpha^\ddagger$-types who can recognize and avoid failed attempts at collaboration with $\alpha^\dagger$-types. And so on, and so forth. For a discussion of collaborative types recognizing their own kind, see Section 5.5 of Newton (2017).

Clearly, when (RG) holds, the payoff from (ii) can only exceed the payoff from (i) if player i free rides on the collaboration of α-types who collaborate without him. That is, somewhat ironically, a necessary condition for invasion by anti-free riding $\alpha^{\varnothing}$-types is that they free ride on the collaborative behavior of α-types.

Consider the prisoner's dilemma at ESS x^*. As non-contribute is a strictly dominant strategy, the payoff $nb-c$ from being the nth coalition member is strictly less than the payoff $(n-1)b$ of free riding on an $n-1$ member coalition. Therefore, $\alpha^{\varnothing}$-types cannot invade.

For the threshold public goods game at ESS x^*, a similar argument holds. $m \leq n-1$ collaborators are chosen to contribute. Whether or not i participates, the good will be provided, so i obtains a higher payoff when he is a ν-type with no chance of being chosen to contribute. Therefore, $\alpha^{\varnothing}$-types cannot invade.

B.4. Robustness to smaller collaborations

Consider the case in which there is some $s_T \in MM(T)$, $\pi_T \in \Pi_T(s_T)$ such that there is a subcoalition $T' \subset T$ that is strictly α-effective for $\pi_{T'}$. That is, if T' operated independently it could assure itself of better payoffs than are assured from the coalitional maxmin s_T for T. To avoid such a possibility, instead of using $MM(T)$, we can define and use $\widetilde{MM}(T) \subseteq MM(T)$,

$$\widetilde{MM}(T) := \left\{ s_T \in S_T : \begin{array}{l} \forall \pi_T \in \Pi_T(s_T), \text{ we have } \pi_T \geq \underline{\pi}_T \text{ and} \\ \forall T' \subseteq T,\ T' \text{ is not strictly } \alpha\text{-effective for } \pi_{T'}. \end{array} \right\} \tag{79}$$

Typically, this does not change much, if anything, in our analysis. For example, our analysis of social dilemmas does not change as a consequence. However, we should be a bit careful as Lemma 1 no longer holds in general. That is, there may exist sets of players $T, T' \subset T$ such that $\widetilde{MM}(T') \neq \varnothing$, $\widetilde{MM}(T) = \varnothing$. If the set of α-type players $N_\alpha = T$, then, in the absence of coalitional maxmin for T, we may wish to allow T' to play some coalitional maxmin profile $s_{T'} \in \widetilde{MM}(T') \subseteq MM(T')$, with the remainder $T \setminus T'$ playing individual maxmin $\underline{s}_{T \setminus T'}$.

Example 3. Consider a 3-player game, $i \in N = \{1,2,3\}$, $S_i = \{1,2,3\}$. For distinct i,j,k with $i = j \mod (3)+1$, let payoffs be as follows.

Case 1. $s_i = s_j = k$.

- $\pi_i(s) = 2$, $\pi_j(s) = 1$.
- If $s_k = k$, then $\pi_k(s) = -4$.
- If $s_k \neq k$, then $\pi_k(s) = 0$.

Case 2. s_i, s_j, s_k are distinct, or $s_i = s_j \in \{i,j\}$. Then, for all $l \in \{i,j,k\}$,

- If $s_l = l$, then $\pi_l(s) = 0$.
- If $s_l \neq l$, then $\pi_l(s) = -5$.

Individual maxmin is $s_i = i$. Coalitional maxmin for $T = \{i,j\}$ is $s_i = s_j = k$. $MM(N)$ consists of s such that $s_i = s_j = k$, $s_k \neq k$. It can be checked that (RG) is satisfied, so Theorem 1 implies that, under our original model, α-types proliferate when rare under any matching protocol.

Now observe that $\widetilde{MM}(N)$ is empty, as for any $s \in MM(N)$, we have $\pi_i(s) = 2$, $\pi_j(s) = 1$, $\pi(k) = 0$. However, $T = \{j,k\}$ is strictly α-effective for payoffs $(1,0)$ as it can guarantee $(2,1)$ by playing $s_j = s_k = i$.

Consider small x. Most groups of three players contain fewer than two α-types, so every player plays individual maxmin and realized payoffs are zero. Therefore, average fitness in the population is close to zero. The question is then whether α-types obtain average payoffs higher than the population average.

If matching is highly positively assortative, so that α-types usually find themselves in groups of three α-types, then their average payoff, when two α-types play coalitional maxmin and the remaining α-type plays individual maxmin, is $(2+1-4)/3 < 0$. Therefore, α-types do not proliferate when rare.

In contrast, when matching is binomial, small x implies that groups with two α-types occur much more frequently than groups with three α-types. The average payoff of α-types in such groups is $(2+1)/2 > 0$, with no negative externalities being imposed on other α-types. Therefore, α-types proliferate when rare.

B.5. Individualistic best responders

Consider a population of ν-types and α-types at ESS x^*. We consider a novel type — the ν^{BR}-type, and ask whether x^* is vulnerable to invasion. Assume that ν^{BR}-types correctly predict how other players in the game will play, then best respond to this prediction. Assume binomial matching, so we can restrict attention to ν^{BR}-types who are matched into groups containing no other ν^{BR}-types.

In the prisoner's dilemma, threshold public goods, minimum effort game and Bertrand competition, the ν^{BR}-type is functionally identical to a ν-type. In the prisoner's dilemma, the best response is always non-contribution. For threshold public goods, either some set of α-types contribute to provide the good or no such set is capable. In either case, the best response is non-contribution. In the

minimum effort game, positive effort is only made when every player is an α-type. Thus, the presence of a ν^{BR}-type implies that other players make zero effort, the best response to which is zero effort. The argument for the Bertrand game is similar.

The case of price competition with imperfect substitutes is more complex. Consider two players ($n = 2$). Theorem 5(ii) implies the unique ESS is $x^* = 1$. When two α-types are matched they play coalitional maxmin $s_\alpha = {}^{D}/_{2(a-2b)}$, giving payoffs of ${}^{D^2}/_{4(a-2b)}$. When a mutant ν^{BR}-type is matched to an α-type, the α-type plays individual maxmin $s_i = {}^{D}/_{2(a-b)}$, to which the ν^{BR}-type best responds with $s_j = {}^{D(2a-b)}/_{4(a-b)^2}$, obtaining a payoff of ${}^{D^2(2a-b)^2}/_{16(a-b)^3}$. This is less than ${}^{D^2}/_{4(a-2b)}$, therefore the mutant ν^{BR}-type cannot invade.

Data availability

No data was used for the research described in the article.

References

Ambrus, A., 2009. Theories of coalitional rationality. J. Econ. Theory 144 (2), 676–695.
Angus, S.D., Newton, J., 2015. Emergence of shared intentionality is coupled to the advance of cumulative culture. PLoS Comput. Biol. 11 (10), e1004587.
Angus, S.D., Newton, J., 2020. Collaboration leads to cooperation on sparse networks. PLoS Comput. Biol. 16 (1), e1007557.
Aumann, R., 1959. Acceptable points in general cooperative n-person games. In: Tucker, A.W., Luce, R.D. (Eds.), Contributions to the Theory of Games IV. Princeton University Press, Princeton, NJ, USA, pp. 287–324.
Aumann, R., Peleg, B., 1960. Von Neumann-Morgenstern solutions to cooperative games without side payments. Bull. Am. Math. Soc. 66.
Bacharach, M., 2006. Beyond Individual Choice: Teams and Frames in Game Theory. Princeton University Press, Princeton, NJ, USA.
Bergstrom, T.C., 1995. On the evolution of altruistic ethical rules for siblings. Am. Econ. Rev., 58–81.
Bernheim, B.D., Peleg, B., Whinston, M.D., 1987. Coalition-proof Nash equilibria I. Concepts. J. Econ. Theory 42 (1), 1–12.
Bratman, M.E., 1992. Shared cooperative activity. Philos. Rev. 101 (2), 327–341.
Butterfill, S., 2012. Joint action and development. Philos. Q. 62 (246), 23–47.
Call, J., 2009. Contrasting the social cognition of humans and nonhuman apes: the shared intentionality hypothesis. Top. Cogn. Sci. 1 (2), 368–379.
Chwe, M.S.-Y., 1994. Farsighted coalitional stability. J. Econ. Theory 63 (2), 299–325.
Cressman, R., 1997. Local stability of smooth selection dynamics for normal form games. Math. Soc. Sci. 34 (1), 1–19.
Dekel, E., Ely, J.C., Yilankaya, O., 2007. Evolution of preferences. Rev. Econ. Stud. 74 (3), 685–704.
Dworczak, P., Pavan, A., 2022. Preparing for the worst but hoping for the best: robust (Bayesian) persuasion. Econometrica 90 (5), 2017–2051.
Farrell, J., Maskin, E., 1989. Renegotiation in repeated games. Games Econ. Behav. 1 (4), 327–360.
Fitch, W.T., 2010. The Evolution of Language:. Approaches to the Evolution of Language. Cambridge University Press, Cambridge, UK.
Frenkel, S., Heller, Y., Teper, R., 2018. The endowment effect as blessing. Int. Econ. Rev. (online first).
Gale, D., Shapley, L.S., 1962. College admissions and the stability of marriage. Am. Math. Mon. 69 (1), 9–15.
Gilbert, M., 1990. Walking together: a paradigmatic social phenomenon. Midwest Stud. Philos. 15 (1), 1–14.
Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with non-unique prior. J. Math. Econ. 18 (2), 141–153.
Gillies, D.B., 1959. Solutions to General Non-zero-Sum Games. Contributions to the Theory of Games, vol. 4 (40), pp. 47–85.
Gold, N., Sugden, R., 2007. Collective intentions and team agency. J. Philos. 104 (3), 109–137.
Güth, W., Kliemt, H., 1998. The indirect evolutionary approach: bridging the gap between rationality and adaptation. Ration. Soc. 10 (3), 377–399.
Heifetz, A., Shannon, C., Spiegel, Y., 2007. What to maximize if you must. J. Econ. Theory 133 (1), 31–57.
Heller, Y., 2014. Overconfidence and diversification. Am. Econ. J. Microecon. 6 (1), 134–153.
Heller, Y., Nehama, I., 2023. Evolutionary foundation for heterogeneity in risk aversion. J. Econ. Theory 208, 105617.
Herings, P.J.-J., Mauleon, A., Vannetelbosch, V., 2009. Farsightedly stable networks. Games Econ. Behav. 67 (2), 526–541.
Jackson, M.O., Watts, A., 2002. On the formation of interaction networks in social coordination games. Games Econ. Behav. 41 (2), 265–291.
Luo, X., Yang, C.-C., 2009. Bayesian coalitional rationalizability. J. Econ. Theory 144 (1), 248–263.
Morris, S., Oyama, D., Takahashi, S., 2023. Implementation via Information Design in Binary-Action Supermodular Games. Available at SSRN 3697335.
Newton, J., 2012. Coalitional stochastic stability. Games Econ. Behav. 75 (2), 842–854.
Newton, J., 2017. Shared intentions: the evolution of collaboration. Games Econ. Behav. 104, 517–534.
Newton, J., 2021a. Conventions under heterogeneous behavioural rules. Rev. Econ. Stud. https://doi.org/10.1093/restud/rdaa063.
Newton, J., 2021b. Corrigendum to “maximality in the farsighted stable set”. Econometrica 89 (4), 18–20.
Newton, J., Angus, S.D., 2015. Coalitions, tipping points and the speed of evolution. J. Econ. Theory 157, 172–187.
Ray, D., Vohra, R., 2015. The farsighted stable set. Econometrica 83 (3), 977–1011.
Ray, D., Vohra, R., 2019. Maximality in the farsighted stable set. Econometrica 87 (5), 1763–1779.
Robson, A.J., 1996. The evolution of attitudes to risk: lottery tickets and relative wealth. Games Econ. Behav. 14 (2), 190–207.
Rommeswinkel, H., 2023. Preference for Verifiability. Waseda Institute for Advanced Study, Waseda University. Mimeo.
Rusch, H., 2019. The evolution of collaboration in symmetric 2×2-games with imperfect recognition of types. Games Econ. Behav. 114, 118–127.
Samuelson, L., 2001. Introduction to the evolution of preferences. J. Econ. Theory 97 (2), 225–230.
Sandholm, W.H., 2010. Local stability under evolutionary game dynamics. Theor. Econ. 5 (1), 27–50.
Searle, J., 1990. Collective intentions and actions. In: Cohen, P.R., Morgan, J., Pollack, M. (Eds.), Intentions in Communication. MIT Press, pp. 401–415.
Taylor, P.D., Jonker, L.B., 1978. Evolutionary stable strategies and game dynamics. Math. Biosci. 40 (1), 145–156.
Tomasello, M., Carpenter, M., 2007. Shared intentionality. Dev. Sci. 10 (1), 121–125.
Tomasello, M., Herrmann, E., 2010. Ape and human cognition what’s the difference? Curr. Dir. Psychol. Sci. 19 (1), 3–8.
Tomasello, M., Rakoczy, H., 2003. What makes human cognition unique? From individual to shared to collective intentionality. Mind Lang. 18 (2), 121–147.
Tuomela, R., Miller, K., 1988. We-intentions. Philos. Stud. 53 (3), 367–389.
Velleman, J.D., 1997. How to share an intention. Philos. Phenomenol. Res. Q. J. 57 (1), 29–50.
Von Neumann, J., Morgenstern, O. Theory of Games and Economic Behavior. 2nd rev.