Games and Economic Behavior 119 (2020) 79–97

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Agency, potential and contagion [☆]

Jonathan Newton [a,*], Damian Sercombe [b]

[a] Institute of Economic Research, Kyoto University, Kyoto, Japan
[b] Einstein Institute of Mathematics, Hebrew University of Jerusalem, Jerusalem, Israel

A R T I C L E   I N F O

Article history:
Received 26 November 2018
Available online 22 October 2019

Keywords:
Agency
Potential
Contagion
Networks

A B S T R A C T

We consider two fundamental forces that can drive the diffusion of an innovation on a network. The first of these forces is potential maximization, a method of aggregating payoff incentives of players under individual agency. Potential maximization is related to the graph theoretic property of close-knittedness (Young, 2011). The second force is collective agency, under which sets of players decide together on whether to adjust their strategies. Collective agency is shown to be related to the graph theoretic property of cohesion (Morris, 2000). We compare the relative strengths of these forces under (i) different payoff specifications in coordination games and (ii) different network structures.

1. Introduction

In 18th century Paris, a boulevardier decides whether to adopt the latest fashion. In 21st century Philadelphia, an economist decides whether to vote in the AEA elections. In Seoul, a university student decides which mobile phone to purchase. In each of these situations, the individuals concerned will be influenced by the choices of those with whom they associate and interact. The diffusion of novel behaviors in such models has been extensively studied. Two important findings have been that insular groups of individuals who interact mainly with one another (i) can be relatively stubborn when it comes to changing their behavior due to external influence (Morris, 2000), but (ii) can exhibit 'autonomy' (Young, 2011) in adopting innovative behaviors ahead of others. Now consider the following situations. Two best friends go to the mall together and decide which next generation games console to buy. A group of colleagues working in a factory discuss and decide whether to join a strike. Some classmates decide which social platform, say WhatsApp or WeChat, to use to organize their study program. These situations are similar to the previous situations but have one important difference – they are instances of collective agency. That is, the individuals concerned are getting together and asking the question *'what should we do?'* rather than asking *'what should I do?'* in isolation. Recent work in developmental psychology suggests that such collaborative thinking and problem solving is a basic human trait, manifesting itself in human infants at ages as early as 14 months (Tomasello, 2014; Tomasello and Rakoczy, 2003, and citations therein). Therefore, the question of whether and how such behavior affects the diffusion of innovation is of great importance.[1]

☆ This work is part of the shared intentions agenda that examines collective agency and social structure. For more, go to http://sharedintentions.net. The authors thank seminar audiences at the University of Tsukuba, University of Kyoto, Monash University, University of Oxford, University of Cambridge, UC Irvine and University of Wisconsin. The manuscript also benefited from conversations with Simon Angus, Jean-Paul Carvalho, Margaret Gilbert, William Sandholm, Ryoji Sawa and Peyton Young. JN is the recipient of a KAKENHI Grant-in-Aid for Research Activity Start-up funded by the Japan Society for the Promotion of Science (Grant Number: 19K20882) and the Research Fund for Young Scientists funded by Kyoto University (Grant Number: 3809930000).

* Corresponding author.
E-mail addresses: newton@kier.kyoto-u.ac.jp (J. Newton), djs213@imperial.ac.uk (D. Sercombe).

[1] Collective agency has been present in modern game theory since the beginning (Neumann, 1928). It underpins concepts such as the Core (Gillies, 1959), Strong Equilibrium (Aumann, 1959), Coalition Proofness (Bernheim et al., 1987; Moreno and Wooders, 1996), Unreliable Team Interaction (Bacharach, 1999),

https://doi.org/10.1016/j.geb.2019.10.007
0899-8256/

In the current paper, we compare the spread of novel behavior due to individual agency, as represented by a potential function, to the spread of novel behavior due to the collective agency of those with strategic complementarities. To do this, we use notions of autonomy. A set of players is said to be *autonomous* if, for whatever reason, they can be expected to adopt a novel behavior regardless of the strategies of players outside of the set. Potential autonomy is defined by means of a potential function (Monderer and Shapley, 1996) that aggregates individual incentives in a manner consistent with individual agency. Local maximizers of potential functions correspond to Nash equilibria (Monderer and Shapley, 1996). Moreover, Ui (2001) shows that if a Nash equilibrium globally maximizes potential, then it is robust to incomplete information in the sense of Kajii and Morris (1997). Global maximizers of potential functions also correspond to stochastically stable states under log-linear dynamics such as the logit choice rule (Blume, 1993). A set of players is *potential autonomous* if, for any fixed strategy profile of players outside of the set, potential is maximized when all the players in the set adopt the novel behavior. Young (2011) showed that potential autonomy depends in a positive way on the graph theoretic property of the close-knittedness of a set of players. Agency autonomy, introduced here, is defined by collective agency. A set of players is *agency autonomous* if, for any fixed strategy profile of players outside of the set, the payoff of every player in the set is increased by collective adoption of the novel behavior. Agency autonomy depends in a positive way on the graph theoretic property of the cohesion of a set of players (Morris, 2000). For the class of symmetric 2×2 coordination games, we give necessary and sufficient conditions for inclusion relations between the set of potential autonomous sets and the set of agency autonomous sets for every possible network of interactions between players. This allows us to classify games according to which form of autonomy leads to wider adoption of the novel behavior. Furthermore, the conditions for our inclusion relations show that the concepts of close-knittedness and cohesion are related via some very natural boundaries between different types of coordination game (e.g. flight to safety, stag hunt, zero off-diagonal, mammoth hunt). Thus, the concept of potential autonomy, which can be used to make precise statements about behavior under low rationality, myopic models of decision making such as the logit choice rule, is intimately linked to agency autonomy, which directly depends on high rationality, collective decision making.[2]

Although these results apply to all networks of interactions, they can be leveraged to make more detailed statements about specific networks. For example, it is shown that the complete graph, the network in which every player interacts with every other player, is more conducive to agency autonomy relative to potential autonomy than any other network of interactions. That is, the complete graph is the network most suited to adoption of novel behavior by means of collective agency relative to individual agency. The opposite result is shown for common classes of random graph, which are amongst the networks most suited to adoption of novel behavior by means of individual agency relative to collective agency.

Members of highly cohesive sets interact relatively little with those outside of the set. Such sets are thus likely to be agency autonomous. That is, cohesion can facilitate the spread of novel behavior. However, the concept of cohesion was introduced by Morris (2000) as something that prevented contagion of a set from the outside. How do these observations relate to one another? It has already been remarked in Newton and Angus (2015) that while cohesion of a set makes it resilient to external contagion, cohesion of its subsets makes it prone to self-contagion via collective agency. Here we generalize that observation, introducing *coalitional cohesion*, a generalization of cohesion. Roughly speaking, the coalitional cohesion of a set of players is increasing in its cohesion, but decreasing in the cohesion of any of its subsets that can exhibit collective agency. A set of players is then immune to contagion, irrespective of the strategies of players outside of the set, if and only if it is sufficiently coalitionally cohesive.

The paper is arranged as follows. Section 2 gives the model and defines potential autonomy and agency autonomy. Section 3 analyzes the relationship between these concepts. Section 4 applies the concepts to trees (graphs without cycles), sparse random graphs and the complete graph. Section 5 provides a characterization of the immunity of groups of players to contagion when the underlying game is any symmetric 2×2 coordination game. Section 6 gives some additional discussion and concludes. Appendix A contains all proofs. Appendix B applies some of the ideas of the paper to the classic karate club network of Zachary (1977).

2. Model

Consider a simple, finite graph $\Gamma = (V, E)$.[3] The vertex set V represents a set of players. The edge set E, consisting of unordered pairs of elements of V, represents connections between players. If two vertices share an edge they are said to be *neighbors*. The number of neighbors of a vertex $i \in V$ is the *degree* of i. For $S \subseteq V$, denote by $d(S)$ the sum of the degrees of vertices in S. Assume that each $i \in V$ has at least one neighbor, so that $d(S) > 0$ for all $S \subseteq V$, $S \neq \emptyset$. For $T, S \subseteq V$, denote by $d(T, S)$ the number of edges $(i, j) \in E$ such that $i \in T$ and $j \in S$. For notational convenience we write $d(\{i\})$ as $d(i)$ and $d(\{i\}, S)$ as $d(i, S)$.

Coalitional Rationalizability (Ambrus, 2009), Coalitional Stochastic Stability (Newton, 2012) and Agency Equilibrium (Newton, 2019). Furthermore, recent work has shown that the ability to participate in collective agency will be evolutionarily selected for in a wide variety of environments (Angus and Newton, 2015; Newton, 2017; Rusch, 2019).

[2] Note that although, aside from Appendix B, this paper does not consider equilibrium or any particular process of strategic updating, the results can be used in such models. For example, if a set of players is agency autonomous and it makes sense in a given setting for that set to exhibit agency, then we can normatively predict the behavior of S under any reasonable equilibrium concept.

[3] A *simple* graph is unweighted, undirected, has no edges from a vertex to itself, and has at most one edge between any pair of vertices.

	A	B
A	$1+\beta$	0
B	$\beta-\alpha$	1

Fig. 1. For each combination of A and B, entries give payoffs for the row player. $\alpha, \beta \in \mathbb{R}$, $\alpha > 0$. This parameterization covers all two player, two strategy symmetric coordination games up to affine transformation of the payoff matrix.

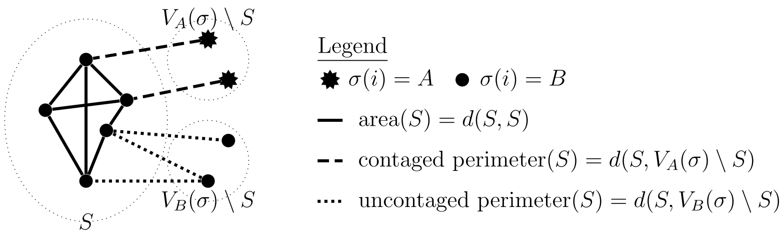


Fig. 2. Illustration of graph theoretic properties. Stars and circles represent players playing A and B respectively. Lines (solid, dashed, dotted) represent edges between neighboring players. In this example, $d(S, S) = 7$, $d(S, V_A(\sigma) \setminus S) = 2$ and $d(S, V_A(\sigma) \setminus S) = 3$.

A strategy profile σ is a function $\sigma : V \to \{A, B\}$ that associates each player with one of two strategies, A or B. Strategy B can be thought of as a status quo and strategy A can be thought of as some novel behavior. As is standard, let σ_S, σ_{-S} denote σ restricted to the domains S and $V \setminus S$ respectively. Let σ^A, σ^B be the strategy profiles such that for all $i \in V$, $\sigma^A(i) = A$, $\sigma^B(i) = B$. Denote by $V_A(\sigma) \subseteq V$ the set of players who play strategy A at profile σ and by $V_B(\sigma) \subseteq V$ the set of players who play strategy B at profile σ. The payoff of a player at profile σ is the sum of his payoffs when he plays his strategy against each of his neighbors on the graph in the game in Fig. 1. Formally, player i's payoff at σ is

$$\pi_i(\sigma) = \begin{cases} (1+\beta)\, d(i, V_A(\sigma)) & \text{if } \sigma(i) = A \\ (\beta-\alpha)\, d(i, V_A(\sigma)) + d(i, V_B(\sigma)) & \text{if } \sigma(i) = B \end{cases}. \tag{2.1}$$

Note that this specification admits an exact potential function (Monderer and Shapley, 1996) given by

$$\mathit{Potential}\,(\sigma) = (1+\alpha)\, d(V_A(\sigma), V_A(\sigma)) + d(V_B(\sigma), V_B(\sigma)). \tag{2.2}$$

The potential function aggregates information from the game in a way that retains information on the incentives of players under individual agency. Specifically, if we adjust the strategy of any single player, the change in his payoff equals the change in the potential function. Note that our payoff specification implies that strategy profile σ^A is the unique global maximizer of (2.2).

However, the potential function does not retain information on the incentives of players under collective agency. For example, a group of players may be able to adjust their strategies together in such a way that every member of the group gains payoff yet potential decreases. To make this clear, it helps, as in Young (2011), to think of $d(S, S)$ as the *area* of S and of $d(S, V \setminus S)$ as the *perimeter* of S. Specifically, for a strategy profile σ, we can think of $d(S, V_A(\sigma) \setminus S)$, the number of neighbors of S who are playing A, as the *contaged perimeter* of S and $d(S, V_B(\sigma) \setminus S)$, the number of neighbors of S who are playing B, as the *uncontaged perimeter* of S (see Fig. 2). Then, from σ such that $S \subseteq V_B(\sigma)$, if we switch all players in S from B to A, the change in potential equals

$$\underbrace{\alpha \cdot \text{area}(S)}_{\text{Internal potential}} + \underbrace{(1+\alpha) \cdot \text{contaged perimeter}(S) - \text{uncontaged perimeter}(S)}_{\text{External potential}},$$

whereas the change in the sum of payoffs of players in S is

$$\underbrace{2\beta \cdot \text{area}(S)}_{\text{Internal coordination}} + \underbrace{(1+\alpha) \cdot \text{contaged perimeter}(S) - \text{uncontaged perimeter}(S)}_{\text{Contagion}}.$$

The contagion effect is the effect of the behavior of players outside of S on the payoffs of S. This effect is perfectly mirrored by the potential function. Where the expressions differ is in the generation of potential and payoffs within S itself. The analysis of individual agency corresponds to S being a singleton, in which case area$(S) = 0$ and neither potential nor payoffs are internally generated.

Potential is independent of β, so from the perspective of individual strategic motivations, the game in Fig. 1 does not change when β is varied. For example, the set of Nash equilibria, including those in mixed strategies, is independent of β. When collective agency is considered, this is clearly no longer the case, and different values of β give very different incentives. If $\beta < 0$, then coordinating on strategy B provides higher payoffs than coordinating on strategy A, so collective agency should work towards retaining strategy B even as differences in potential promote the adoption of strategy A. If

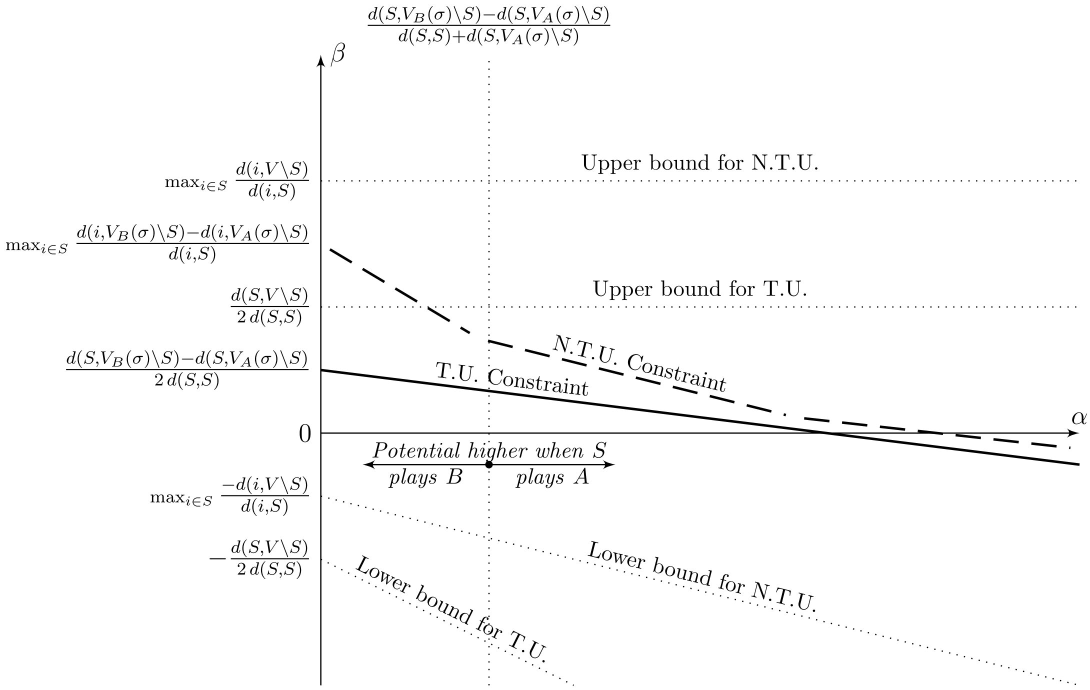


Fig. 3. Fixing strategies of players outside S, if all players in S switch from B to A, the sum of their payoffs (resp. the payoff of each player in S) increases if and only if (α, β) lies above TU Constraint (resp. NTU Constraint) which is weakly decreasing and linear (resp. weakly decreasing, piecewise linear, convex). Upper (lower) bounds for these constraints are attained when all players outside S play B (A). Areas in which such a switch increases (decreases) potential are illustrated.

$\beta > 0$, then agency and potential work in the same direction to promote the spread of strategy A. The expressions above indicate that potential change and the sums of payoff changes are precisely aligned when $\beta = \alpha/2$. However, this does not consider that an increase in the sum of payoffs of a set of players is not enough to guarantee that every player in the set gains. Collective choice by a set of players will be constrained by those who do worst out of any anticipated change.

This last point is illustrated in Fig. 3, which uses the expressions given above to illustrate the conditions on α and β under which a switch by S from B to A will (i) increase the sum of the payoffs of players in S and so be a rational coalitional move under transferable utility (TU) constraints, (ii) increase the payoff of every player in S and so be a rational coalitional move under non-transferable utility (NTU) constraints, (iii) increase potential.

2.1. Autonomy

To compare individual and collective agency we shall use ideas of autonomy. A set of players S is autonomous if there is some reasonable expectation that players in the set will come to play A regardless of the choices of players outside of S. There are different ways in which such a reasonable expectation might arise. We examine two of them. Potential autonomy considers the behavior of a group as determined by an aggregation of individual incentives in a way that is consistent with individual agency.

Definition 1. $S \subseteq V$ is potential autonomous if, for all σ such that $\sigma_S \neq \sigma_S^A$,

$$\text{Potential}\,(\sigma_S^A, \sigma_{-S}) > \text{Potential}\,(\sigma).$$

That is, S is potential autonomous if, for any strategies played by players outside of S, a higher potential is attained when players in S all play A than when they play any other strategies. It has been proven that, under asynchronous log-linear learning dynamics, potential maximizing strategy profiles are observed more frequently than other strategy profiles in the long run (Blume, 1993). Moreover, if S is potential autonomous then the hitting time for strategy profiles such that $\sigma(i) = A$ for all $i \in S$ can be bounded above independently of the rest of the network. Convergence to A is fast for potential autonomous sets (Young, 2011).

From a static perspective, given a set S, arbitrarily fix the strategies of players outside of S and consider the induced game with player set S. If S is potential autonomous, then σ_S^A maximizes potential in the induced game, so by Theorem 3 of Ui (2001), σ_S^A is a Nash equilibrium which is robust to incomplete information in the sense of Kajii and Morris (1997).

Young (2011) shows that potential autonomy depends on the graph theoretic property of close-knittedness, which measures how well integrated each subset of a group of players is with the rest of the group. The *close-knittedness* of a set $S \subseteq V$ is given by

$$CK(S) := \min_{S' \subseteq S} \frac{d(S', S)}{d(S')}.$$

Remark 1 *(Young, 2011). S is potential autonomous if and only if $CK(S) > \frac{1}{2+\alpha}$.*

Our second concept of autonomy is agency autonomy. Agency autonomy does not require aggregation of incentives but instead considers collective agency. If, given the choice, all members of S would like to play A, conditional on others in S also playing A, regardless of what players in $V \setminus S$ do, we say that S is agency autonomous.

Definition 2. $S \subseteq V$ is agency autonomous if, for all σ, for all $i \in S$,

$$\pi_i(\sigma_S^A, \sigma_{-S}) > \pi_i(\sigma_S^B, \sigma_{-S}).$$

Similarly to potential autonomy, agency autonomy can be related to a graph theoretic concept, cohesion (Morris, 2000), which measures how well integrated into S is the least well integrated member of S. The *cohesion* of a set S is given by

$$Co(S) := \min_{i \in S} \frac{d(i, S)}{d(i)}.$$

As we consider coordination games, the inequality in the definition of agency autonomy is hardest to satisfy when $\sigma = \sigma^B$. Then, for all $i \in S$, we have $\pi_i(\sigma_S^A, \sigma_{-S}^B) = (1+\beta)d(i, S)$ and $\pi_i(\sigma_S^B, \sigma_{-S}^B) = d(i)$, so the condition in the definition of agency autonomy becomes $(1+\beta)d(i, S) > d(i)$. Rearranging and taking the minimum over all $i \in S$, we have

Remark 2. S is agency autonomous if and only if $Co(S) > \frac{1}{1+\beta}$.[4]

Remarks 1 and 2 are illustrated in Fig. 4, the shaded areas in the Figure showing which values of α and β correspond to S being potential autonomous and agency autonomous respectively.

The reader may note that potential autonomy is defined by individual agency but depends on close-knittedness, which measures the integration of the least well integrated group within S. In contrast, agency autonomy is defined by collective agency but depends on cohesion, which measure the integration of the least well integrated individual within S. This contrast arises because agency autonomy does not aggregate individual incentives, so that each individual in S holds a veto with regard to collective changes in strategy. If we were to instead consider the change in the sum of payoffs of S, then $Co(S)$ would be replaced by ${}^{2d(S,S)}/_{d(S)}$ in Remark 2.

3. Relations between potential autonomy and agency autonomy

In order to examine the relationship between potential autonomy and agency autonomy, we shall first establish some foundational results that link close-knittedness and cohesion. That $Co(.)$ is bounded below by $CK(.)$ is remarked in Young (2011). We show that it is also bounded above by $2CK(.)$.

Lemma 1. *For given $\Gamma = (V, E)$, $S \subseteq V$, $0 \le CK(S) \le Co(S) \le 2CK(S) \le 1$.*

We say that a set of players S is homogeneous if no subset of S is less well integrated with S than S is with itself. That is, S is *homogeneous* if

$$CK(S) = \frac{d(S, S)}{d(S)}.$$

By Remark 1, to check whether a homogeneous set is potential autonomous only requires us to check whether ${}^{d(S,S)}/_{d(S)} > {}^{1}/_{(2+\alpha)}$.

[4] As $Co(S) \le 1$, Remark 2 implies that agency autonomy is only possible when $\beta > 0$. It is possible to define a concept of reverse-agency autonomy under which sets wish to play B, regardless of the choices of those outside of the set. This is only possible for $\beta < 0$. The upward sloping boundaries (in α-β space) for inclusion relations between potential autonomy and agency autonomy that we obtain in Propositions 1,2,3 are instead downwards sloping boundaries for inclusion relations between potential autonomy and reverse-agency autonomy. The analysis repeats much of what we do for agency autonomy and so is omitted from the current exposition.

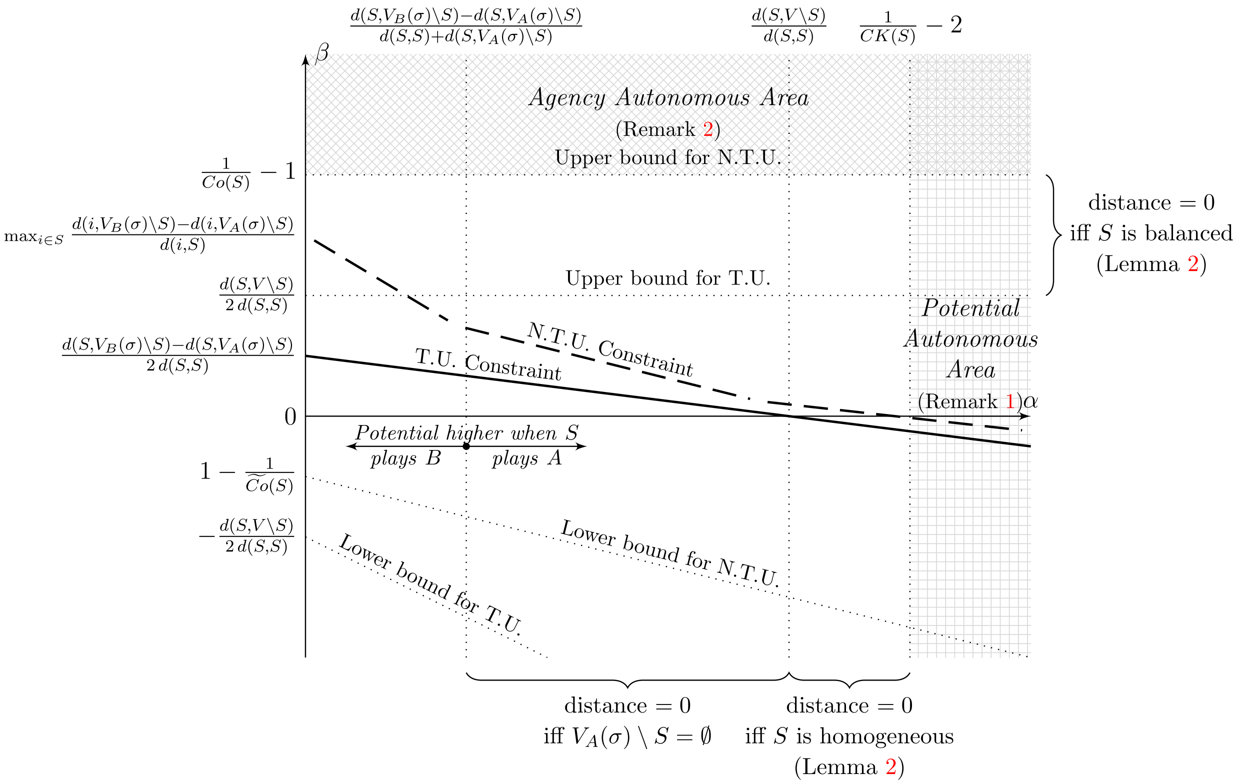


Fig. 4. For a given set of players S, shaded areas indicate parameter values under which S is potential autonomous and agency autonomous. The relationship between potential maximization, the strategies of players outside of S, homogeneity, close-knittedness and potential autonomy can be seen reading the diagram horizontally. The relationship between collective strategy choice, balancedness, cohesion and agency autonomy can be seen reading the diagram vertically.

Define *anti-cohesion* as how well integrated into S is the most well integrated member of S.

$$\widetilde{Co}(S)=\max_{i\in S}\frac{d(i,S)}{d(i)}.$$

When cohesion equals anti-cohesion, $Co(S)=\widetilde{Co}(S)$, then every member of S is equally well integrated into S and we say that S is *balanced*. It follows immediately from this definition that S is balanced if and only if ${}^{d(i,S)}/_{d(i)}$ is constant across all $i\in S$. For sets with a given proportion of within-set interactions, homogeneous sets maximize close-knittedness and balanced sets maximize cohesion.

Lemma 2. *$CK(S)\leq\frac{d(S,S)}{d(S)}$, with equality if and only if S is homogeneous. $Co(S)\leq\frac{2d(S,S)}{d(S)}$, with equality if and only if S is balanced.*

The bounds in Lemma 2 (illustrated in Fig. 4) can then be used to show that when S is balanced it must also be homogeneous, and that the ratio of cohesion to close-knittedness attains its upper bound. This is useful, as balance is defined by equality between ${}^{d(i,S)}/_{d(i)}$ for $|S|$ possible $i\in S$, whereas homogeneity is defined by $2^{|S|}$ possible $S'\subseteq S$, so balance will usually be easier to check than homogeneity.

Lemma 3. *For given $\Gamma=(V,E)$, if $S\subseteq V$ is balanced, then S is homogeneous and $Co(S)=2\,CK(S)$.*

We are now in a position to compare potential autonomy and agency autonomy. The first proposition concerns a condition on α and β under which, for every graph $\Gamma=(V,E)$, every potential autonomous set is agency autonomous.

Proposition 1. *$\beta\geq 1+\alpha$ if and only if for all $\Gamma=(V,E)$, every potential autonomous $S\subseteq V$ is also agency autonomous.*

That is to say, when $\beta\geq 1+\alpha$, collective agency can contribute at least as much, possibly more, to the spread of novel behavior (strategy A) as can individual agency combined with the perturbations necessary to attain local potential maximizing profiles. Considering our games in α-β space (Fig. 5), we see that $\beta\geq 1+\alpha$ corresponds to the area bounded

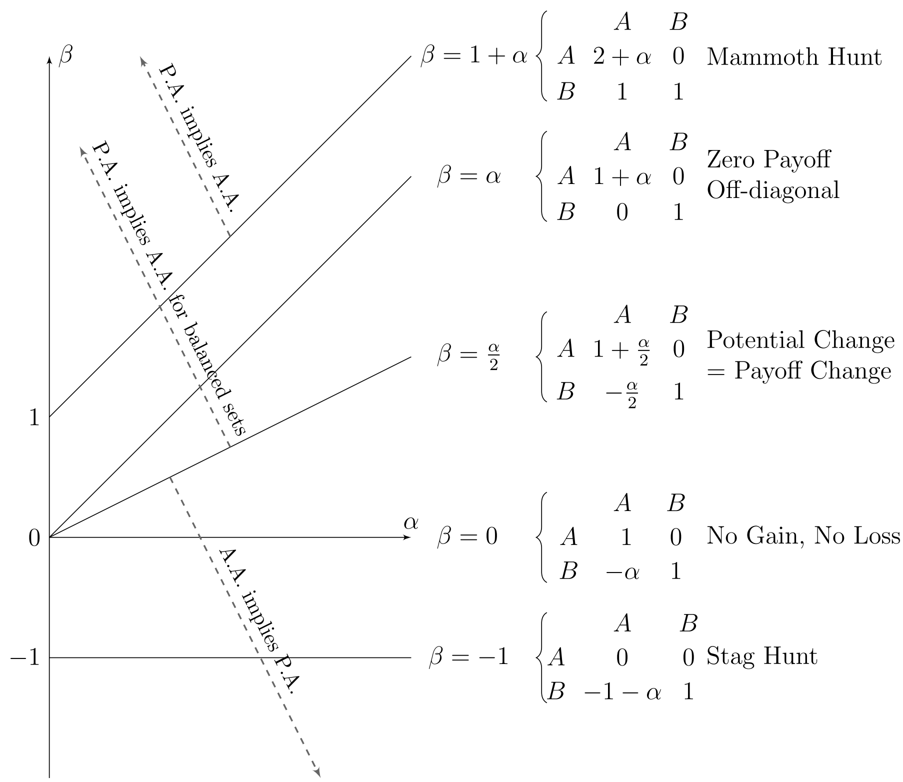


Fig. 5. Propositions 1, 2, 3 illustrated in α-β space. The ordinal ranking of payoffs in the game changes at $\beta = 1+\alpha$, $\beta = \alpha$, $\beta = 0$, $\beta = -1$. At $\beta = \alpha/2$, if σ_S changes, then the consequent change in the sum of the payoffs of players in S exactly equals the change in potential.

below by the 'mammoth hunt' game, which is similar to a 'stag hunt', but with the payoff for stag-stag increased to make stag-stag the potential maximizing profile and hence the risk-dominant Nash equilibrium. In Fig. 5, in the area below the mammoth hunt the ordinal payoff ranking of the game changes, so Proposition 1 tells us that the area of Fig. 5 in which potential autonomy implies agency autonomy corresponds exactly to the area in which the payoff ranking of our game has (A, A) preferred to (B, A) to (B, B) to (A, B).

To understand why the bound in Proposition 1 is $\beta = 1+\alpha$, consider a set $T \subseteq V$ that is sufficiently well integrated to be both potential autonomous and agency autonomous. Let the set $S = T \cup \{i\}$ consist of T together with a single player $i \notin T$. We can think of i as the least well integrated member of S. Now, given that T is potential autonomous, a necessary and sufficient condition for S to be potential autonomous is that, starting from $(\sigma_T^A, \sigma_{-T}^B)$, potential increases when player i switches to strategy A. From (2.2), this condition is

$$(1+\alpha)\, d(i, T) - d(i, V \setminus T) = (2+\alpha)\, d(i, T) - d(i) > 0. \tag{3.1}$$

Given that T is agency autonomous, we have (see Footnote 4) that $\beta > 0$. This implies that strategy pair (A, A) gives a higher payoff than (A, B), so players in T will never be deterred from switching to A by the prospect of player i switching with them. Therefore, a necessary and sufficient condition for S to be agency autonomous is that, starting from σ^B, the payoff of player i increases when all of the players in S switch to A. From (2.1), this condition is

$$(1+\beta)\, d(i, T) - d(i) > 0. \tag{3.2}$$

Comparing (3.1) and (3.2), we see that the conditions are equivalent only when $\beta = 1+\alpha$. In all other cases, one of the conditions is easier to satisfy than the other.

In the above discussion, the threshold $\beta = 1+\alpha$ arises because, by considering general values of $d(i, T)$, $d(i, V \setminus T)$, we allow player i to be arbitrarily well integrated in S, and thus $Co(S)$ to be arbitrarily low. At the other extreme, we can mandate that S be balanced so that every member of S is equally well integrated in S. As a consequence of balance, the individual incentives of each player in S are perfectly aligned with the goal of maximizing the sum of payoffs over all of the players in S. As we found in Section 2, this corresponds to a threshold of $\beta = \alpha/2$.

Proposition 2. *$\beta \geq \alpha/2$ if and only if for all $\Gamma = (V, E)$, every balanced, potential autonomous $S \subseteq V$ is also agency autonomous.*

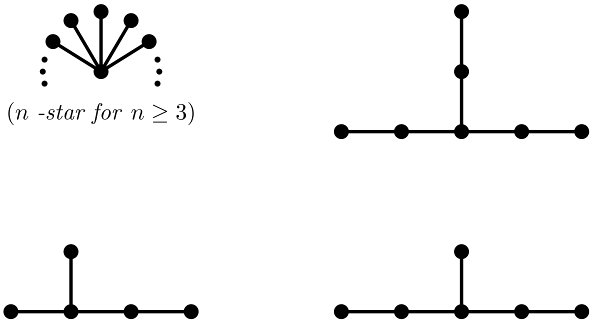


Fig. 6. Exceptional trees.

For balanced sets, Proposition 2 expands the implications of Proposition 1 to a larger class of games, such as when $\beta = \alpha$ and the game is a coordination game with zero payoffs off the main diagonal.

Next consider the reverse problem, to find conditions under which agency autonomy implies potential autonomy. The bounding case will be when the condition for agency autonomy (Remark 2) is as easy as possible to satisfy relative to the condition for potential autonomy (Remark 1). This case arises when S is balanced. To see this, consider that $Co(S) = 2CK(S)$ for balanced S (Lemma 3), but $Co(S) \leq 2CK(S)$ for all S (Lemma 1), so for given α, β, the condition for agency autonomy is easiest to satisfy relative to the condition for potential autonomy when S is balanced. So we must consider balanced S. As discussed before Proposition 2, this makes the problem simple, and we once again have $\beta = \alpha/2$ as a bound.

Proposition 3. *$\beta \leq \alpha/2$ if and only if for all $\Gamma = (V, E)$, every agency autonomous $S \subseteq V$ is also potential autonomous.*

So, for low values of β, any behavioral process that works towards potential maximization can contribute at least as much, possibly more, to the spread of novel behavior (strategy A) as can collective agency. Proposition 3 is trivially true for $\beta \leq 0$, as in this case there exist no agency autonomous sets. This class includes $\beta = 0$, where neither Nash equilibrium of the game Pareto dominates the other, and $\beta = -1$, the stag hunt. For the latter case, we would actually expect any collective agency to work against the adoption of strategy A, as when $\beta = -1$, any individual's payoff from strategy A is independent of whether or not others play A.

So we have completed a simple comparison of potential autonomy and agency autonomy. Both forms of autonomy would seem to be important, giving contrasting ways of considering the behavior of groups. Potential autonomy aggregates individual incentives in a way that is consistent with individual agency. Agency autonomy leaves individual incentives as they are, but aggregates agency. We now move to consider these concepts for a particular class of graphs, in the process proving some results that are of independent graph theoretic interest.

4. Examples

4.1. Trees

In this section we relax the assumption that $\Gamma = (V, E)$ be finite and require only that there be a finite upper bound on the degree of vertices in Γ. For finite $S \subseteq V$, $CK(S)$, $Co(S)$, homogeneity, balancedness and agency autonomy remain well defined. We extend the definition of potential autonomy by saying that finite $S \subseteq V$ is potential autonomous if it is potential autonomous on any finite subgraph of Γ that includes S and all adjacent edges $(i, j) \in E$, $i \in S$.

A *tree* is a connected graph that contains no cycles. An *n-regular* tree is the unique (up to isomorphism) tree where each vertex has degree n, where n is a positive integer. Note that any n-regular tree with $n \geq 2$ must be infinite. For given $\Gamma = (V, E)$ and $S \subseteq V$, we say that S is *connected* if the induced subgraph with vertex set S and edge set $E_S = \{(i, j) \in E : i, j \in S\}$ is connected. For an n-regular tree, any finite, connected S, $|S| \geq 2$, will always include at least one vertex with precisely one neighbor in S. Therefore, $Co(S) = 1/n$. To find $CK(S)$, we first prove a result on homogeneity.

An *n-quasiregular* tree is a connected component of any graph constructed by removing up to n edges from the n-regular tree. In particular, all except for at most n vertices of an n-quasiregular tree have degree n, and no vertices have a degree larger than n. A *quasiregular* tree is an n-quasiregular tree for some $n \in \mathbb{N}$. The *n-star* is the tree with $n + 1$ vertices where one vertex is adjacent to all of the others.

Proposition 4. *Let $\Gamma = (V, E)$ be a tree. Every finite, connected $S \subseteq V$ is homogeneous if and only if Γ is quasiregular or is one of the exceptional trees illustrated in Fig. 6.*

One implication of this result is that every finite, connected S on an n-regular tree is homogeneous and

$$CK(S) = \frac{d(S,S)}{d(S)} = \frac{|S|-1}{|S|n}.$$

So $CK(S)$ is increasing in $|S|$. For $|S|=2$, $CK(S)={}^1/_{2n}$ and as $|S|\to\infty$, $CK(S)\to{}^1/_n$. Given that, for $|S|\geq 2$, $Co(S)={}^1/_n$, these are, according to the bounds in Lemma 1, the lowest and highest values that $CK(S)$ could take. Notice that $Co(S)=2\,CK(S)$ only when $|S|=1$ or 2, so by Lemma 3, if $|S|>2$, then S cannot be balanced. We have

Lemma 4. *Let $\Gamma=(V,E)$ be an n-regular tree. Let $S\subseteq V$ be finite and connected. Then S is balanced if and only if either $|S|=1$ or $|S|=2$.*

Although Lemma 4 concerns n-regular trees,[5] it can be used to prove the analogue of Proposition 4 for the balance property on general trees.

Proposition 5. *Let $\Gamma=(V,E)$ be a tree. Every finite, connected $S\subseteq V$ is balanced if and only if Γ is the (unique) tree with $|V|=2$.*

Turning our attention once more to n-regular trees, Remarks 1, 2 and our expressions for $Co(S)$, $CK(S)$ combine to show that for potential autonomy to imply agency autonomy for given finite, connected S, $|S|\geq 2$, we require

$$\beta \geq 1+\alpha-\frac{2+\alpha}{|S|}. \tag{4.1}$$

As the right hand side of (4.1) is increasing in $|S|$ and approaches $1+\alpha$ as $|S|\to\infty$, it follows that if we wish potential autonomy to imply agency autonomy for any finite, connected S, we require $\beta\geq 1+\alpha$. That is, for n-regular trees, the bound in Proposition 1 is attained. In contrast, if $|S|=2$, then, by Lemma 4, S is balanced, so by Proposition 2, we only require $\beta\geq{}^\alpha/_2$.

4.2. Random graphs

An *ensemble* of random graphs is a probability measure over a set of graphs. Popular random graph ensembles include Erdös-Renyi graphs, random regular graphs and the Configuration Model (for definitions, see Bollobás, 2001). For the named ensembles, if in addition to randomly choosing a graph, we randomly choose a vertex within the graph, then the probability that the neighborhood of the chosen vertex is a *random tree* approaches one as the number of vertices increases. For a precise definition of this convergence, see Dembo and Montanari (2010), whose definition of random tree we adapt as follows.

Let $P=\{P_n : n\in\mathbb{N}\}$ be a probability measure on vertex degrees, with finite, positive first moment, and denote by $\rho_n = nP_n/\sum_{l=0}^{\infty} lP_l$ its size-biased version. Let $P_0=P_1=0$, so that every vertex has degree at least two. Let $\top(P,m)$ denote the ensemble of random trees (V,E) generated as follows. Start from a *root* vertex $i_1\in V$. Choose an integer n according to P, then add edges between i_1 and n new vertices that we add to V. These n vertices constitute the next *generation*. Continue recursively as follows. For each vertex in the previous generation, generate an integer n independently according to ρ, and connect the vertex to $n-1$ new vertices. Repeat m times.

Let $\mathbb{P}_m$ give probabilities over pairs (Γ,S), where $\Gamma=(V,E)$ and $S\subseteq V$, $|S|=m$. Let $\mathbb{P}_m$ be determined by the following rule. First, randomly choose Γ according to $\top(P,m)$. Denote the root vertex by $i_1\in V$ and let $S_1=\{i_1\}$. Then, iterating for $r=2,\dots,m$, uniformly at random choose $j\in S_{r-1}$ such that j has at least one neighbor who is not in S_{r-1}. Uniformly at random choose a neighbor of j, say k, who is not in S_{r-1} and let $S_r=S_{r-1}\cup\{k\}$. Thus $\mathbb{P}_m$ gives probabilities over sets S, the neighborhoods of which are given by random trees.

Consider $\mathbb{P}_m$ as m becomes large. Either there exists a maximum degree of a vertex, say $\hat{n}$, in which case, with high probability, $Co(S)$ and $CK(S)$ approach ${}^1/_{\hat{n}}$; or there exists no maximum degree, in which case $Co(S)$ and $CK(S)$ approach zero. In either case, with high probability, $Co(S)$ and $CK(S)$ take similar values, and we have the following lemma.

Lemma 5. *For all $\varepsilon>0$,*

$$\lim_{m\to\infty}\mathbb{P}_m\Big[\big\{(\Gamma,S) : Co(S)-CK(S)>\varepsilon\big\}\Big]=0.$$

When $Co(S)\approx CK(S)$, if $\beta<1+\alpha$, then the condition in Remark 1 is easier to satisfy than the condition in Remark 2. Consequently, we can use Lemma 5 to show that if $\beta<1+\alpha$, then, as S becomes large, the probability of S being agency autonomous but not potential autonomous approaches zero.

[5] For general trees, finite, connected $S\subseteq V$, $|S|\geq 2$, there will always be some $i\in S$ such that $d(i,S)=1$. Let $n:=d(i)$. By definition of balancedness, we then have that S is balanced if and only if ${}^{d(j,S)}/_{d(j)}={}^1/_n$ for all $j\in S$. If the tree is regular, then $d(j)=n$ for all $j\in S$, so this condition reduces to $d(j,S)=1$ for all $j\in S$. This is only possible when $|S|=2$ and is thus an alternative proof of Lemma 4.

Proposition 6. *For given Γ, let $\mathcal{T}_{\alpha,\beta,\Gamma}$ be the set of all $S \subseteq V$ that are agency autonomous but not potential autonomous. If $\beta < 1+\alpha$, then*

$$\lim_{m\to\infty} \mathbb{P}_m\Big[\big\{(\Gamma, S) : S \in \mathcal{T}_{\alpha,\beta,\Gamma}\big\}\Big] = 0,$$

So for large, connected S, $\beta = 1+\alpha$ provides a tight bound between potential autonomy implying agency autonomy (Proposition 1) and vice versa (Proposition 6). Thus (random) trees are amongst the graphs least susceptible to contagion driven by collective agency as compared to contagion driven by differences in potential.

4.3. Complete graphs

$\Gamma = (V, E)$ is *complete* if $\{i, j\} \in E$ for all $i, j \in V$, $i \neq j$. For all $S \subseteq V$, $i \in S$, we have $d(i, S) = |S| - 1$, $d(i) = |V| - 1$. Therefore, for all $S \subseteq V$, $i \in S$,

$$\frac{d(i,S)}{d(i)} = \frac{|S|-1}{|V|-1} = Co(S),$$

so S is balanced. Furthermore, the relationship between completeness and balance goes in both directions. The complete graph is the unique graph for which every (connected) $S \subseteq V$ is balanced.

Lemma 6. *Γ is complete if and only if every connected $S \subseteq V$ (alternatively, every $S \subseteq V$) is balanced.*

For balanced S, $\beta = \alpha/2$ is a tight bound between potential autonomy implying agency autonomy (Proposition 2) and vice versa (Proposition 3). Thus the complete graph is amongst the graphs most susceptible to contagion driven by collective agency as compared to contagion driven by differences in potential.

5. Immunity to contagion

Let Ω, a set of finite subsets of V, be a set of *feasible coalitions*. Assume that if S is a feasible coalition ($S \in \Omega$), then any subset of S is also a feasible coalition ($T \subseteq S \implies T \in \Omega$). A set S is immune to contagion by strategy A if, no matter the strategies of players outside of S, given that all $i \in S$ are playing B, no feasible coalition within S will want to switch to A.

Definition 3. $S \subseteq V$ is immune to contagion if, for all σ, there does not exist $T \subseteq S$, $T \in \Omega$ such that for all $i \in T$,

$$\pi_i(\sigma_T^A, \sigma_{S\setminus T}^B, \sigma_{V\setminus S}) - \pi_i(\sigma_S^B, \sigma_{V\setminus S}) > 0. \tag{5.1}$$

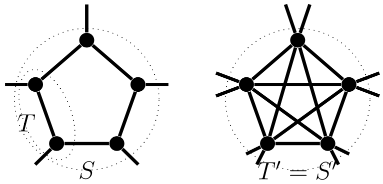
Consider the binding case when all players outside of S are already playing strategy A. Whether a feasible coalition $T \subset S$ will switch to strategy A depends on the incentives of the member of T who has the least to gain from the switch. This will be a player in T who is well integrated in S but not too well integrated in T. If this player is sufficiently integrated in S relative to T then he will veto any switch by T to strategy A. For S to be immune to contagion, all $T \subseteq S$, $T \in \Omega$ must contain such a pivotal player. Of these pivotal players, there will exist one who is least integrated in S relative to T. The feasible coalition in which this player is pivotal would be the first domino to fall. The level of integration of this player in S relative to T is the *coalitional cohesion* of S.

$$CoCo(S,\alpha,\beta) := \min_{\substack{T\subseteq S\\ T\in\Omega}} \max_{i\in T} \frac{d(i,S)}{d(i)} - \frac{1+\beta}{2+\alpha}\frac{d(i,T)}{d(i)}.$$

Proposition 7. *S is immune to contagion if and only if $CoCo(S,\alpha,\beta) \geq \frac{1+\alpha}{2+\alpha}$.*

Roughly speaking, S must be cohesive enough to avoid contagion from the outside, but not contain subgroups which are sufficiently cohesive themselves to wish to switch to A together (see Fig. 7). Note that if Ω is just the set of singletons, then $CoCo(S,\alpha,\beta) = Co(S)$ and we have $Co(S) \geq \frac{1+\alpha}{2+\alpha}$ as the condition in Proposition 7, which is effectively the threshold in Proposition 1 of Morris (2000). Similarly, for the class of 'panic' games ($\beta \leq -1$) bounded above by the stag hunt in Fig. 5, the minimum in $CoCo(S,\alpha,\beta)$ is attained when T is a singleton and we again have that coalitional cohesion equals cohesion.

Note that $CoCo(S,\alpha,\beta)$ is, in general, not independent of the game parameters α and β. However, it is indeed independent for several salient values of β, such as $\beta = -1$ (stag hunt), $\beta = \alpha/2$ (potential change = change in sum of payoffs), and $\beta = 1+\alpha$ (mammoth hunt). This characterization extends Proposition 3 of Newton and Angus (2015), which effectively deals with the case $\beta = \alpha$ (zero payoff off-diagonal). When $\beta = 1+\alpha$ (mammoth hunt), the weightings of $d(i,S)$ and $d(i,T)$



$$Co(S) = \tfrac{2}{3} = Co(S')$$

$$CoCo(S, \alpha, \beta) = \tfrac{2}{3} - \tfrac{1+\beta}{2+\alpha}\tfrac{1}{3}$$

$$CoCo(S', \alpha, \beta) = \tfrac{2}{3} - \tfrac{1+\beta}{2+\alpha}\tfrac{2}{3}$$

Fig. 7. Let $\Gamma = (V, E)$ contain the subgraphs on vertex sets S and S' shown above. Let Ω be the set of *cliques*, sets $T \subseteq V$ such that every member of T is a neighbor of every other member of T. If $\beta > -1$, then values for cohesion and coalitional cohesion are as given above. Note that although S and S' are equally cohesive, more integrated coalitions are possible within S' than within S ($^{d(i,T')}/_{d(i)} = {}^2/_3$ for $i \in T'$, whereas $^{d(i,T)}/_{d(i)} = {}^1/_3$ for $i \in T$). Consequently, S' is less coalitionally cohesive than S. If $\beta \leq -1$, then coalitional cohesion equals cohesion as discussed in the text.

in the expression for $CoCo(S, \alpha, \beta)$ are identical. This is the case of a status quo strategy that gives a constant payoff, so that Proposition 7 in this case is effectively Lemma 1 of Reich (2016).

Unions of immune sets are not necessarily immune. Consider S, S' that are immune to contagion. Let $i \in S$, $j \in S'$ and $T = \{i, j\} \in \Omega$. If i and j switch from B to A together, each will gain additional payoff β from the edge that they share. If i had already switched to A, then, by switching to A, j would gain additional payoff $1 + \alpha$ from his edge shared with i. Therefore, if $\beta > 1 + \alpha$, the incentive for j to participate in a joint switch together with i is greater than the incentive for j to switch after i has already switched. Consequently, it is possible that $S \cup S'$ is not immune to contagion. If $\beta \leq 1 + \alpha$, then this logic is reversed, so that immunity of S, S' implies immunity of $S \cup S'$.

Proposition 8. *Let α be a rational number. Then $\beta \leq 1 + \alpha$ if and only if for all $\Gamma = (V, E)$, Ω, any $S, S' \subseteq V$ which are immune to contagion have a union $S \cup S'$ which is also immune to contagion.*

It is possible to build on Proposition 8 and the intuition behind it to give results on dynamic processes of strategic updating. For example, when V is finite and $\beta \leq 1 + \alpha$, there exists a largest set that is immune to contagion. One might then expect that, starting from σ^B, a myopic coalitional updating rule would eventually converge to a profile at which all players outside of this set play A and all players inside the set play B. The approach of the current paper has been to abstain from discussion of any particular rule for strategy updating, so aside from a brief discussion of Zachary's karate club in Appendix B, we leave such analysis to other work.[6] The reader seeking guidance on how to build such a model is referred to Newton (2018).

6. Discussion

We end the paper with a brief discussion of further relationships with the existing literature and prospects for extending our analysis.

6.1. The philosophy of conventions

In his classic work on conventions, Lewis (1969) restricts attention to games in which, if we fix the strategy of any given player, then that player prefers that other players also play that same strategy. In our notation, this is equivalent to $-1 \leq \beta \leq 1 + \alpha$. Gilbert (1981) later persuasively argued that this restriction was unnecessary. In the current paper, Proposition 1 shows that the upper threshold is important to the comparison of potential maximization and collective agency, and Proposition 8 shows that it is important to immunity to contagion in the presence of coalitions.

6.2. Coalitional potential functions

Sawa (2014) defines a *coalitional potential function* as a function such that any change in σ_S leads to a change in potential equal to the sum of the changes in the payoffs of players in S. The examples in the cited paper are broad and economically relevant (e.g. exchange economies), but they rely on an absence of externalities. That is, when a coalition S updates its strategies, the payoffs of players outside of S are unaffected. This is not the case in the current paper, where, as discussed in Section 2, a coalitional potential function of this type is impossible unless $\beta = {}^{\alpha}/_2$.

[6] For example, there is a literature on dynamics that lead to the play of either risk dominant or payoff dominant strategies. Risk dominance has a tendency to succeed under perturbed individualistic best response dynamics (Blume, 2003; Peski, 2010; Norman, 2009), whereas rules that select payoff dominance tend to be more elaborate (Pradelski and Young, 2012; Marden et al., 2014). The assumption that $\alpha > 0$ implies that A is the risk dominant action. If $\beta < 0$ then B is the payoff dominant action. When this is the case, collective agency will work towards coalitions of players choosing B and in favor of payoff dominance.

6.3. No potential function

Our analytical framework can clearly be applied to any game that admits a potential function. However, the use of potential functions does impose limitations. One such limitation is evident in the current context if we consider a directed graph of interactions instead of an undirected one. Let $V = \{i, j\}$ and let there be a single directed edge from player i to player j so that player j obtains payoff from this interaction but player i does not. The resulting game does not admit a potential function. Of course, the limitations of potential functions do not make them unimportant. Aside from the popular classes of games that do admit potential functions, it is also the case that for any game to have a pure strategy Nash equilibrium, it must have some aspect of a potential game embedded in its payoff structure (Candogan et al., 2011).[7]

6.4. Afterword

Before writing this paper, it was not obvious to the authors that interesting connections would exist between the concepts that we consider. After reading, we hope that such connections, and the reasons for them, are clear and transparent. Potential maximization (aggregation of individual incentives) is linked to close-knittedness via payoffs, which is linked to cohesion via graph theory, which is linked to collective agency (aggregation of agency) via payoffs. The elucidation of this connection should be of use to others seeking to bridge the individual and the collective.

Appendix A. Proofs

Proof of Lemma 1.

$$
\begin{aligned}
2\,CK(S) &= 2\min_{S'\subseteq S}\frac{d(S',S)}{d(S')} = 2\min_{S'\subseteq S}\frac{\sum_{i\in S'} d(i, S\setminus S') + \frac{1}{2}d(i,S')}{\sum_{i\in S'} d(i)} \\
&\geq 2\min_{S'\subseteq S}\frac{\sum_{i\in S'} \frac{1}{2}d(i, S\setminus S') + \frac{1}{2}d(i,S')}{\sum_{i\in S'} d(i)} = \min_{S'\subseteq S}\frac{\sum_{i\in S'} d(i,S)}{\sum_{i\in S'} d(i)} \\
&\underbrace{\geq}_{\substack{\text{by mediant}\\ \text{inequality}}} \min_{S'\subseteq S}\min_{i\in S'}\frac{d(i,S)}{d(i)} = \underbrace{\min_{i\in S}\frac{d(i,S)}{d(i)}}_{=Co(S)} \geq \min_{S'\subseteq S}\frac{d(S',S)}{d(S')} = CK(S). \qquad \square
\end{aligned}
$$

Proof of Lemma 2. The result for $CK(S)$ follows by definition. For $Co(S)$,

$$
Co(S) = \min_{i\in S}\frac{d(i,S)}{d(i)} \underbrace{\leq}_{\substack{\text{by mediant}\\ \text{inequality}}} \frac{\sum_{j\in S} d(j,S)}{\sum_{j\in S} d(j)} = \frac{2\,d(S,S)}{d(S)},
$$

and

$$
\begin{aligned}
S \text{ balanced} &\Leftrightarrow \min_{i\in S}\frac{d(i,S)}{d(i)} = \max_{i\in S}\frac{d(i,S)}{d(i)} \Leftrightarrow \forall i,j\in S,\ \frac{d(i,S)}{d(i)} = \frac{d(j,S)}{d(j)} \\
&\underbrace{\Leftrightarrow}_{\substack{\text{by mediant}\\ \text{inequality}}} \forall i\in S,\ \frac{d(i,S)}{d(i)} = \frac{\sum_{j\in S} d(j,S)}{\sum_{j\in S} d(j)} = \frac{2\,d(S,S)}{d(S)} \Leftrightarrow Co(S) = \frac{2\,d(S,S)}{d(S)}. \qquad \square
\end{aligned}
$$

Proof of Lemma 3.

$$
CK(S) \underbrace{\geq}_{\text{by Lemma 1}} \frac{1}{2}Co(S) \underbrace{=}_{\text{by Lemma 2}} \frac{d(S,S)}{d(S)} \geq \min_{S'\subseteq S}\frac{d(S',S)}{d(S')} = CK(S) \qquad \square
$$

Definition 4. Let $\mathcal{P}(\Gamma,\alpha)$ denote the set of potential autonomous sets and $\mathcal{A}(\Gamma,\alpha)$ denote the set of agency autonomous sets.

Proof of Proposition 1. If $\beta \geq 1+\alpha$, $S \in \mathcal{P}(\Gamma,\alpha)$, then

$$
Co(S) \underbrace{\geq}_{\text{by Lemma 1}} CK(S) \underbrace{>}_{\text{by Remark 1}} \frac{1}{2+\alpha} \underbrace{\geq}_{\text{as } \beta\geq 1+\alpha} \frac{1}{1+\beta}.
$$

[7] Of course, many games have no pure Nash equilibrium. In particular, Goldberg et al. (1968) show that if random payoffs are generated for an m by n two-person noncooperative game, the probability that a realization has a Nash equilibrium in pure strategies approaches $1 - 1/e$ as $m, n \to \infty$.

so by Remark 2, we have that $S \in \mathcal{A}(\Gamma, \beta)$.

To show that the bound on β is tight, for any $\beta < 1+\alpha$ we construct a graph $\Gamma = (V, E)$ which includes a set $S \subseteq V$ such that $S \in \mathcal{P}(\Gamma, \alpha)$ but $S \notin \mathcal{A}(\Gamma, \beta)$. Consider Γ that includes a clique $S \subseteq V$. Let $i \in S$ be the only vertex in S which has any neighbors outside of S. Denote $T = S \setminus \{i\}$.

Note that $Co(T) = \widetilde{Co}(T) = {}^{(|T|-1)}/_{|T|}$, so T is balanced and, by Lemma 3, $Co(T) = 2CK(T)$. So as $|T| \to \infty$, $Co(T) \to 1$ and $CK(T) \to {}^1/_2$, implying that for large enough $|T|$, T is agency autonomous and potential autonomous.

If T is potential autonomous, then S is potential autonomous if and only if, from $(\sigma^A_T, \sigma^B_{-T})$, potential increases when player i switches to A. Using (2.2), this condition is

$$|T|(2+\alpha) - d(i) > 0. \tag{A.1}$$

If T is agency autonomous, then S is agency autonomous if and only if, from σ^B, the payoff of player i increases when S switches to A. Using (2.1), this condition is

$$|T|(1+\beta) - d(i) > 0. \tag{A.2}$$

Therefore, for any $\beta < 1+\alpha$, $|T|$ and $d(i)$ can be chosen such that $|T|$ is large enough that T is potential autonomous, and the ratio of $|T|$ to $d(i)$ is such that (A.1) holds but (A.2) does not. So S is potential autonomous but not agency autonomous. □

Proof of Proposition 2. If $\beta \geq {}^\alpha/_2$, $S \in \mathcal{P}(\Gamma, \alpha)$, S is balanced, then

$$Co(S) \underbrace{=}_{\substack{\text{by balance}\\ \text{and Lemma 3}}} 2CK(S) \underbrace{>}_{\text{by Remark 1}} \frac{2}{2+\alpha} \underbrace{\geq}_{\text{by } \beta \geq \alpha/2} \frac{1}{1+\beta},$$

so by Remark 2, we have $S \in \mathcal{A}(\Gamma, \beta)$.

Now assume $\beta < {}^\alpha/_2$. Let $\Gamma = (V, E)$ be complete. It follows that the induced subgraph on any set of vertices S is also complete. Therefore, for any $S \subseteq V$, $\frac{d(i,S)}{d(i)} = \frac{|S|-1}{|V|-1} = Co(S)$ for all $i \in S$, so S is balanced.

If $\beta \leq 0$, then by Remark 2 and Lemma 1, $\mathcal{A}(\Gamma, \beta) = \emptyset$. However, as any $S \subseteq V$ is balanced, V is balanced and

$$CK(V) \underbrace{=}_{\substack{\text{by balance}\\ \text{and Lemma 3}}} \frac{1}{2}Co(V) \underbrace{=}_{\substack{\text{by balance}\\ \text{and Lemma 2}}} \frac{d(V,V)}{d(V)} = \frac{1}{2} \underbrace{>}_{\text{as } \alpha > 0} \frac{1}{2+\alpha},$$

hence, by Remark 1, we have that $V \in \mathcal{P}(\Gamma, \alpha)$ and the inclusion fails.

If $0 < \beta < {}^\alpha/_2$, then we have $\frac{2}{2+\alpha} < \frac{1}{1+\beta}$. Choose $|V|$, $S \subseteq V$, such that $\frac{2}{2+\alpha} < Co(S) < \frac{1}{1+\beta}$. By Remark 2 we have that S is not agency autonomous. As S is balanced, by Lemma 3 we have $Co(S) = 2CK(S)$. Therefore $CK(S) > \frac{1}{2+\alpha}$ and by Remark 1 we have that S is potential autonomous. □

Proof of Proposition 3. If $\beta \leq {}^\alpha/_2$, $S \in \mathcal{A}(\Gamma, \beta)$, then

$$CK(S) \underbrace{\geq}_{\text{by Lemma 1}} \frac{1}{2}Co(S) \underbrace{>}_{\text{by Remark 2}} \frac{1}{2}\frac{1}{1+\beta} \underbrace{\geq}_{\text{by } \beta \leq \alpha/2} \frac{1}{2+\alpha}$$

so $S \in \mathcal{P}(\Gamma, \alpha)$ by Remark 1.

If $\beta > {}^\alpha/_2$, similarly to the proof of Proposition 2, let $\Gamma = (V, E)$ be complete so that any $S \subseteq V$ is balanced and $Co(S) = \frac{|S|-1}{|V|-1}$. Choose $|V|$, $S \subseteq V$, such that $\frac{2}{2+\alpha} > Co(S) > \frac{1}{1+\beta}$. By Remark 2 we have that S is agency autonomous. As S is balanced, by Lemma 3 we have $Co(S) = 2CK(S)$. Therefore $CK(S) < \frac{1}{2+\alpha}$ and by Remark 1 we have that S is not potential autonomous. □

Proof of Proposition 4. Let n be the maximal degree of any vertex in Γ. By the assumption of the model that every vertex has at least one neighbor, we have $n > 0$. Observe that an n-quasiregular tree is always infinite if $n > 2$.

($\Longrightarrow$) Assume that Γ is an infinite tree that is not quasiregular. We construct a pair of subsets $S' \subset S$ of the vertex set of Γ as follows, depending on whether Case (i) or Case (ii) holds.

Case (i). There exist only finitely many vertices $\{v_1, \ldots, v_N\}$ in Γ of degree strictly less than n. By definition of n-quasiregularity, we must have $d(\{v_1, \ldots, v_N\}) < nN - n$. Let $\boldsymbol{P}$ be the smallest subtree in Γ containing all vertices in the set $\{v_1, \ldots, v_N\}$. Let v be any vertex in $\Gamma \setminus \boldsymbol{P}$ that is a neighbor of some vertex in $\boldsymbol{P}$. Let $\boldsymbol{S}$ be the smallest subtree in Γ containing both $\boldsymbol{P}$ and v.

Case (ii). There exist infinitely many vertices in Γ of degree strictly less than n. Let v be any vertex in Γ of maximal degree n. Since Γ is infinite, there exists at least one connected component C_0 of $\Gamma \setminus \{v\}$ that contains infinitely many vertices of degree strictly less than n. Take any $n+1$ of these vertices in C_0, denote them $\{w_1, ..., w_{n+1}\}$. Let $\boldsymbol{S}$ be the smallest subtree in Γ containing all vertices in the set $\{v, w_1, ..., w_{n+1}\}$.

In both cases, denote the vertex set of $\boldsymbol{S}$ by S, and let $k := |S|$. By construction, we have the inequality $d(S) < nk - n$. Let S' be the singleton set consisting only of v. Recall that $v \in S$ has maximal degree n and is adjacent to precisely one other vertex in $\boldsymbol{S}$. Then we have

$$\frac{d(S,S)}{d(S)} > \frac{k-1}{nk-n} = \frac{1}{n} = \frac{d(S',S)}{d(S')}$$

and so S is not homogeneous.

Now assume that Γ is a finite tree that is not quasiregular and that every finite, connected $S \subseteq V$ is homogeneous. Let v be any vertex in Γ of maximal degree n, with adjacent edges $\{e_1, ..., e_n\}$. Consider the set of connected components $\{\boldsymbol{C}_1, ..., \boldsymbol{C}_n\}$ of the graph $\Gamma \smallsetminus \{v, e_1, ..., e_n\}$. For all $i \in \{1, ..., n\}$, let C_i denote the vertex set of $\boldsymbol{C}_i$ and let $c_i := |C_i|$. If $d(C_i) < nc_i - n$ for some $i \in \{1, ..., n\}$ then the result follows from the same arguments as in Case (ii) where Γ is infinite, taking S to be $C_i \cup \{v\}$ and S' to be $\{v\}$.

Now consider the case of $d(C_i) \geq nc_i - n$ for all $i \in \{1, ..., n\}$. This greatly limits the possibilities for what Γ can be. Since each $\boldsymbol{C}_i$ is a finite tree, at least one vertex in each $\boldsymbol{C}_i$ must have degree 1 in Γ. Hence, for all $i \in \{1, ..., n\}$, $d(C_i)$ must equal either $nc_i - n$ or $nc_i - (n-1)$. What possible finite trees $\boldsymbol{C}_i$ and Γ satisfy this condition?

If $n = 3$ then the only valid possibility for such a $\boldsymbol{C}_i$ is either a single vertex or the unique tree on two vertices, and hence the only possibilities for Γ are exceptional trees as displayed in Fig. 6. If $n > 3$ then the only valid possibility for such a $\boldsymbol{C}_i$ is a single vertex, and hence the only possibility for Γ is the n-star graph (also an exceptional tree displayed in Fig. 6). Since all finite trees with $n < 3$ are quasiregular (they are all subtrees of the 2-regular tree) we may ignore the case $n < 3$.

($\Longleftarrow$) Assume Γ is quasiregular (so it must be n-quasiregular). Let $\boldsymbol{T}$ be any finite subtree of Γ, with vertex set T, and let T' be any proper subset of T. Denote $l := |T|$ and $l' := |T'|$. Observe that $d(T,T) = l-1$ and $d(T',T) \geq l'$ (since $\boldsymbol{T}$ is a tree and $l' < l$). Moreover, $d(T) \geq nl - n$ (by definition of n-quasiregularity) and $d(T') \leq l'n$. Summarizing, we have

$$\frac{d(T,T)}{d(T)} \leq \frac{l-1}{nl-n} = \frac{1}{n} = \frac{l'}{l'n} \leq \frac{d(T',T)}{d(T')}$$

and so T is homogeneous.

It remains to check that every possible coalition in each exceptional tree in Fig. 6 is indeed homogeneous. We check this for the n-star, and leave it as an easy exercise to the reader to check the remaining exceptional trees.

Let $\boldsymbol{U}$, with vertex set U, be any subtree of the n-star, where $n \geq 3$. If U consists of only a single vertex then it is trivially homogeneous, so without loss of generality assume that U contains the unique vertex v of degree n and $u \geq 0$ vertices of degree 1. Let U' be any subset of U. Say U' contains $u' \leq u$ vertices of degree 1. If $v \notin U'$ then $d(U',U) = d(U') = u'$ and if $v \in U'$ then $d(U',U) = d(U,U) = u$ and $d(U') = u' + n$. In either case, we have

$$\frac{d(U,U)}{d(U)} = \frac{u}{u+n} \leq \frac{d(U',U)}{d(U')}$$

and so U is homogeneous. $\square$

Proof of Lemma 4. Immediate from discussion in the text. $\square$

Proof of Proposition 5. Let $\Gamma = (V,E)$ be a tree such that every finite, connected $S \subseteq V$ is balanced. Let i and j be any two vertices in V and let their respective degrees be k_1 and k_2. Let S be the smallest subtree in Γ containing both vertices i and j. Since S is balanced, we have ${}^1\!/_{k_1} = {}^{d(i,S)}\!/_{d(i)} = {}^{d(j,S)}\!/_{d(j)} = {}^1\!/_{k_2}$, therefore $k_1 = k_2 =: n$. So Γ must be an n-regular tree. However, by Lemma 4, any finite, connected $S \subseteq V$, $|S| > 2$, in an n-regular tree is not balanced. Hence Γ must be the tree on two vertices. $\square$

Proof of Lemma 5. Consider $P_2 = 1$, so $\rho_2 = 1$. In this case, the ensemble $\top(P,m)$ only includes a single graph, the line $\Gamma = (V,E)$. By construction, $\mathbb{P}_m$ then puts all probability on connected sets $S \subset V$ that do not include the two vertices with degree 1 at opposite ends of the line Γ. $Co(S) = {}^1\!/_2$ and $CK(S) = {}^{(m-1)}\!/_{(2m)}$. Therefore, $Co(S) - CK(S) \to 0$ as $m \to \infty$ and we are done.

For the rest of the proof, assume that $P_2 < 1$, so $\rho_2 < 1$. For given (Γ, S), denote

$$\mathcal{L}(S) = \{i \in S : d(i,S) = 1\}.$$

As a first step, we show that for any $r_0 \in \mathbb{N}$,

$$\lim_{m\to\infty} \mathbb{P}_m\Big[\Big\{(\Gamma,S) : |\mathcal{L}(S)| < r_0\Big\}\Big] = 0. \tag{A.3}$$

To see this, consider any given sequence $\{S_1, \ldots, S_k\}$, $k < m$, constructed as described in the main body of the text prior to the statement of the lemma. Note that $\mathcal{L}(\cdot)$ is weakly increasing on such sequences and $|\mathcal{L}(S_1)| = 0$, $|\mathcal{L}(S_2)| = 2$. It must be that $|\mathcal{L}(S_k)| = r$ for some $r \in \mathbb{N}$.

Consider continuations of $S_1, \ldots, S_k$ to obtain $\{S_1, \ldots, S_k, \ldots, S_m\}$, with $m - k$ an even number. Consider $\{S_1, \ldots, S_m\}$ for which $|\mathcal{L}(S_m)| = r$. This implies that $|\mathcal{L}(S_t)| = r$ for all $k \le t \le m$. Let $\bar{m} = {}^{(k+m)}/_2$. There are two cases to consider.

Case A. There does not exist t, $k \le t \le \bar{m}$, $i \in S_t$, such that $i \notin \mathcal{L}(S_t)$, $d(i, S_t) \neq d(i)$.

For $k \le t \le \bar{m} - 1$, under Case A, vertices in $\mathcal{L}(S_t)$ are the only vertices in S_t that have any neighbors in Γ that are outside of S_t. Therefore, S_{t+1} is obtained from adding a neighbor, say j, of some $i \in \mathcal{L}(S_t)$, to S_t. Consequently, $i \notin \mathcal{L}(S_{t+1})$ and $j \in \mathcal{L}(S_{t+1})$. Furthermore, it must be that $d(i) = 2$ or otherwise Case A would be contradicted (as i would then have more than 1, but fewer than $d(i)$ neighbors).

Consequently, if $S_{t+1} = S_t \cup \{j\}$ and $d(j) > 2$, then $j \in \mathcal{L}(S_{t+1})$ but S_{t+1} cannot be further extended by adding a neighbor of j, as doing so would contradict Case A. As $|\mathcal{L}(S_t)| = r$ on $\{S_k, \ldots, S_{\bar{m}}\}$, the addition of j with $d(j) > 2$ can therefore occur at most r times before it becomes impossible to extend S_t without contradicting either $|\mathcal{L}(S_t)| = r$ or Case A.

As a consequence of the above, the probability of Case A is bounded above by the probability that r or fewer vertices of degree strictly greater than 2 are added on $\{S_k, \ldots, S_{\bar{m}}\}$. The probability under $\top(P, m)$ that any vertex added to S_t, $t \ge 1$, has degree strictly greater than 2 is $(1 - \rho_2)$, so the probability of r or fewer such additions on $\{S_k, \ldots, S_{\bar{m}}\}$ is

$$\sum_{\mu=0}^{r} \binom{\bar{m} - k}{\mu} (\rho_2)^{\bar{m}-k-\mu} (1 - \rho_2)^{\mu},$$

which approaches zero as $\bar{m} \to \infty$, which, by definition of $\bar{m}$, occurs as $m \to \infty$.

Case B. There exists $\tilde{t}$, $k \le \tilde{t} \le \bar{m}$, $i \in S_{\tilde{t}}$ such that $i \notin \mathcal{L}(S_{\tilde{t}})$, $d(i, S_{\tilde{t}}) \neq d(i)$.

Consider $\tilde{t} \le t \le m - 1$. If $i \in S_t$ and a neighbor of i, say j, is chosen so that $S_{t+1} = S_t \cup \{j\}$, then it must be that $d(i, S_t) \neq d(i)$. If $i \notin \mathcal{L}(S_t)$, then $\mathcal{L}(S_{t+1}) = \mathcal{L}(S_t) \cup \{j\}$, so $|\mathcal{L}(S_{t+1})| = |\mathcal{L}(S_t)| + 1$, contradicting $|\mathcal{L}(S_t)| = r$ for all $t = k, \ldots, m$. So it must be that for each t, $S_{t+1} = S_t \cup \{j\}$ for some neighbor j of some $i \in \mathcal{L}(S_t)$. As $|\mathcal{L}(S_t)| = r$, the probability of choosing $i \in \mathcal{L}(S_t)$ rather than some $i \notin \mathcal{L}(S_t)$ (which is possible, given Case B) is no greater than ${}^{r}/_{(r+1)}$. This must happen at $t = \tilde{t}, \ldots, m - 1$, which occurs with a probability no greater than

$$\left(\frac{r}{r+1}\right)^{m-\tilde{t}} \le \left(\frac{r}{r+1}\right)^{m-\bar{m}} = \left(\frac{r}{r+1}\right)^{\frac{m-k}{2}},$$

which approaches zero as $m \to \infty$.

So, from any S_k, $|\mathcal{L}(S_k)| = r$, the probability of $|\mathcal{L}(S_m)| \le r$ can be made arbitrarily small by increasing m. Consequently, for any $r_0 \in \mathbb{N}$, we can make $\mathbb{P}_m\big[\big\{(\Gamma, S) : |\mathcal{L}(S)| < r_0\big\}\big]$ arbitrarily small by increasing m and we have (A.3).

Consider a given tree $\Gamma = (V, E)$, with root vertex i_1, and vertex $j \in U \subset V$ such that $j \neq i_1$. It will be shown that the probability that $\mathcal{L}(S_m) = U$ is independent of $d(j)$. First note that if U includes some l such that the unique path in Γ from l to i_1 passes through j, then it is impossible that $j \in \mathcal{L}(S_m)$ and thus impossible that $U = \mathcal{L}(S_m)$. Assuming that this is not the case, if we create a new graph by altering the branches of Γ that emerge from j in any direction other than towards i_1 while maintaining $d(j) \ge 2$, then the probability of $\mathcal{L}(S_m) = U$ remains the same. This is because the altered part of the graph has no effect on the construction of $\{S_1, \ldots\}$ unless and until (i) S_t is reached such that $j \in \mathcal{L}(S_t)$; and (ii) j is randomly chosen so that S_{t+1} will be the union of S_t with some neighbor of j. However, if this occurs, then $j \notin \mathcal{L}(S_{t+1})$, regardless of the value of $d(j)$.

In summary, the probability of $U = \mathcal{L}(S)$ is independent of $d(j)$ for $j \in U$, $j \neq i_1$. That is, knowing that a vertex j is in $\mathcal{L}(S)$ does not tell us anything more about $d(j)$ than what we already know, that it is chosen according to the distribution ρ. In particular, the probability that all vertices in $\mathcal{L}(S) \setminus \{i_1\}$ have degree less than $\hat{n}$, conditional on $\mathcal{L}(S)$ containing at least r vertices, is bounded:

$$\mathbb{P}_m\Big[\Big\{(\Gamma, S) : d(j) < \hat{n} \text{ for all } j \in \mathcal{L}(S)\Big\} \Big| |\mathcal{L}(S)| \ge r\Big] \le \left(\sum_{n<\hat{n}} \rho_n\right)^{r-1}, \tag{A.4}$$

where the bound is constructed by multiplying independent probabilities of $r - 1$ vertices in $\mathcal{L}(S) \setminus \{i_1\}$ having degree less than $\hat{n}$.

Let $\hat{n}$ be such that $\sum_{n<\hat{n}} \rho_n < 1$. For any given $\delta > 0$, it is possible to choose $\tilde{r}$ large enough such that $\left(\sum_{n<\hat{n}} \rho_n\right)^{\tilde{r}-1} < {}^{\delta}/_2$. Furthermore, (A.3) implies that it is possible to choose $\tilde{m}$ such that $\mathbb{P}_{\tilde{m}}\big[\big\{(\Gamma, S) : |\mathcal{L}(S)| < \tilde{r}\big\}\big] < {}^{\delta}/_2$. Therefore

$$\begin{aligned}
&\mathbb{P}_{\tilde{m}}\big[\,\big\{\,(\Gamma,S): d(j)\geq \hat{n} \text{ for some } j\in\mathcal{L}(S)\,\big\}\,\big] \\
&\geq \underbrace{\mathbb{P}_{\tilde{m}}\big[\,\big\{\,(\Gamma,S): |\mathcal{L}(S)|\geq \tilde{r}\,\big\}\,\big]}_{>\left(1-\frac{\delta}{2}\right) \text{ by choice of } \tilde{m}}\cdot \underbrace{\mathbb{P}_{\tilde{m}}\big[\,\big\{\,(\Gamma,S): d(j)\geq \hat{n} \text{ for some } j\in\mathcal{L}(S)\,\big\}\,\big|\,|\mathcal{L}(S)|\geq \tilde{r}\,\big]}_{>\left(1-\frac{\delta}{2}\right) \text{ by (A.4)}} \\
&> \left(1-\frac{\delta}{2}\right)^2 > 1-\delta.
\end{aligned} \tag{A.5}$$

As, for any $\delta > 0$, (A.5) holds for large enough $\tilde{m}$, we have

$$\lim_{m\to\infty} \mathbb{P}_m\big[\,\big\{\,(\Gamma,S): d(j)\geq \hat{n} \text{ for some } j\in\mathcal{L}(S)\,\big\}\,\big] = 1. \tag{A.6}$$

If $j \in \mathcal{L}(S)$, $d(j) \geq \hat{n}$, then $Co(S) \leq {}^1/_{\hat{n}}$. So, from (A.6),

$$\lim_{m\to\infty} \mathbb{P}_m\left[\left\{(\Gamma,S): Co(S)\leq \frac{1}{\hat{n}}\right\}\right] = 1. \tag{A.7}$$

Now consider $CK(S)$. By Lemma 1, $CK(S) \leq Co(S)$. If there exists no maximal n such that $P_n > 0$, then for any $\varepsilon > 0$, we can choose $\hat{n}$ such that ${}^1/_{\hat{n}} < \varepsilon$, so by (A.7),

$$\lim_{m\to\infty} \mathbb{P}_m\big[\,\big\{\,(\Gamma,S): Co(S)-CK(S)>\varepsilon\,\big\}\,\big] = 0, \tag{A.8}$$

proving the lemma.

If there exists maximal n such that $P_n > 0$, let $\hat{n}$ take this value. Recall that $CK(S) = \min_{S'\subseteq S} \frac{d(S',S)}{d(S')}$. If $S = S'$, $d(S', S) = m-1$ and $d(S) \leq m\hat{n}$. If $S' \subseteq S$, $S' \neq S$, then $d(S', S) \geq |S'|$, $d(S') \leq |S'|\hat{n}$. In either case, $\frac{d(S',S)}{d(S')} \geq \frac{(m-1)}{m\hat{n}}$. This lower bound on $CK(S)$ approaches ${}^1/_{\hat{n}}$ as $m \to \infty$, which together with $CK(S) \leq Co(S)$ (by Lemma 1) and (A.7), implies (A.8) and we are done. □

Proof of Proposition 6. Assuming $\beta < 1 + \alpha$, let

$$\varepsilon = \frac{1}{1+\beta} - \frac{1}{2+\alpha} > 0.$$

If (Γ, S) is such that S is agency autonomous but not potential autonomous, then by Remarks 2, 1 respectively, $Co(S) > \frac{1}{1+\beta}$ and $CK(S) \leq \frac{1}{2+\alpha}$, so

$$Co(S) - CK(S) > \frac{1}{1+\beta} - \frac{1}{2+\alpha} = \varepsilon,$$

but, by Lemma 5, the probability of such an (Γ, S) pair approaches zero as $m \to \infty$. □

Proof of Lemma 6. If Γ is complete, then balancedness of all $S \subseteq V$ follows from the argument immediately prior to the statement of the Lemma. To prove the reverse, assume that all connected $S \subseteq V$ are balanced. Choose arbitrary $i \in V$. Γ connected implies that there exists $j \in V$ such that $\{i, j\} \in E$. $S = \{i, j\}$ is connected and $d(i, S) = d(j, S) = 1$. Balancedness of S implies that $\frac{d(i,S)}{d(i)} = \frac{d(j,S)}{d(j)}$, so we have $d(i) = d(j)$. This argument applies to any pair of vertices that share an edge in Γ, so as Γ is connected, all vertices in V have the same degree.

For $S = \{i, j\}$, the induced subgraph of Γ on S is complete. For connected S, $|S| = m$, $m \leq |V|$, we show by induction that the induced subgraph on S is complete. For given S, $|S| = m$, choose $i_0 \in S$ such that $T = S \setminus \{i_0\}$ is connected. By induction on m, it must be that $\{j, k\} \in E$ for all $\{j, k\} \subseteq T$, so $d(j, T) = |T| - 1$ for all $j \in T$. By definition of T, at least one vertex, say $i_1 \in T$, is a neighbor of i_0, therefore $d(i_1, S) = |T| = |S| - 1$. For all $j \in S$, balancedness of S implies that $\frac{d(j,S)}{d(j)} = \frac{d(i_1,S)}{d(i_1)}$, and as all vertices in V have the same degree, $d(j) = d(i_1)$ so $d(j, S) = d(i_1, S) = |S| - 1$. Therefore, the induced subgraph of Γ on S is complete. As Γ is connected, this holds for $S = V$, therefore Γ is complete. □

Proof of Proposition 7. The condition in Definition 3 can be written as

$$\min_{\substack{T\subseteq S\\ T\in\Omega}} \max_{i\in T} \pi_i(\sigma_S^B, \sigma_{V\setminus S}^A) - \pi_i(\sigma_T^A, \sigma_{S\setminus T}^B, \sigma_{V\setminus S}^A) \geq 0.$$

Substituting payoffs this becomes

$$\min_{\substack{T\subseteq S\\ T\in\Omega}} \max_{i\in T} (\beta-\alpha)d(i, V\setminus S) + d(i,S) - (1+\beta)(d(i,T) + d(i, V\setminus S)) \geq 0,$$

which rearranges to give the required condition. □

Proof of Proposition 8. Consider $\beta \leq 1+\alpha$. Let S', S'' be immune to contagion. Let $S = S' \cup S''$. Let $T \subseteq S$ attain the minimum in the definition of $CoCo(S, \alpha, \beta)$. Assume, without loss of generality, that $T' := S' \cap T$ is nonempty. Then

$$CoCo(S,\alpha,\beta) = \max_{i \in T} \frac{d(i,S)}{d(i)} - \frac{1+\beta}{2+\alpha}\frac{d(i,T)}{d(i)} \underbrace{\geq}_{\text{by } T' \subseteq T} \max_{i \in T'} \frac{d(i,S)}{d(i)} - \frac{1+\beta}{2+\alpha}\frac{d(i,T)}{d(i)}$$

$$= \max_{i \in T'} \frac{d(i,S')}{d(i)} - \frac{1+\beta}{2+\alpha}\frac{d(i,T')}{d(i)} + \underbrace{\frac{d(i,S\setminus S')}{d(i)} - \frac{1+\beta}{2+\alpha}\frac{d(i,T\setminus T')}{d(i)}}_{\geq 0 \text{ by } \beta \leq 1+\alpha \text{ and } T\setminus T' \subseteq S \setminus S'}$$

$$\geq \max_{i \in T'} \frac{d(i,S')}{d(i)} - \frac{1+\beta}{2+\alpha}\frac{d(i,T')}{d(i)} \underbrace{\geq}_{\substack{\text{by defn} \\ \text{of } Coco(.)}} CoCo(S',\alpha,\beta) \underbrace{\geq}_{\substack{\text{by immunity of } S' \\ \text{and Proposition 7}}} \frac{1+\alpha}{2+\alpha},$$

so by Proposition 7, $S = S' \cup S''$ is immune to contagion.

Now consider $\beta > 1+\alpha$. Let $\Gamma = (V, E)$ contain two cliques S' and S'', each on $k+1$ vertices. Let each vertex in $S' \cup S''$ have degree $k+l$ and let there be precisely one edge $e = \{i, j\}$ joining S' and S'' (that is, $d(S', S'') = 1$). Assume that the restriction of the set Ω of feasible coalitions to $S' \cup S''$ is the union of $\{e\}$ with the set of singletons. Then

$$CoCo(S',\alpha,\beta) \underbrace{=}_{\substack{\text{as all feasible} \\ \text{coalitions in } S \\ \text{are singletons}}} Co(S') \underbrace{=}_{\substack{\text{by defn} \\ \text{of } Co(.)}} \frac{k}{k+l}.$$

Similarly, we have that $CoCo(S'',\alpha,\beta) = {}^{k}/_{(k+l)}$. Also

$$CoCo(S' \cup S'',\alpha,\beta) \underbrace{\leq}_{\substack{\text{by defn} \\ \text{of } Coco(.)}} \frac{d(i,S' \cup S'')}{d(i)} - \frac{1+\beta}{2+\alpha}\frac{d(i,e)}{d(i)}$$

$$= \frac{k+1}{k+l} - \frac{1+\beta}{2+\alpha}\frac{1}{k+l} \underbrace{<}_{\text{by } \beta > 1+\alpha} \frac{k}{k+l}.$$

As α is rational, we can choose k and l such that ${}^{(1+\alpha)}/_{(2+\alpha)} = {}^{k}/_{(k+l)}$. Consequently,

$$CoCo(S' \cup S'',\alpha,\beta) < \frac{1+\alpha}{2+\alpha} = CoCo(S',\alpha,\beta) = CoCo(S'',\alpha,\beta),$$

and Proposition 7 implies that S' and S'' are immune to contagion but $S' \cup S''$ is not. □

Appendix B. Zachary's Karate Club

Here we consider our ideas in relation to Zachary's Karate Club - a social network famous in the literature. The original paper by Zachary (1977) followed the members of a Karate Club in the early 1970s that split into two resulting subclubs, a new club led by the instructor (member 1) and the remnants of the old club led by the administrator (member 34).

We shall first model the observed split in Zachary's Karate Club as a single coalitional move, using the theory of the current paper to find the values of β for which this move is rational. Next, we consider a family of dynamic updating processes in which a coalition first breaks away from the club, following which individual members sequentially choose to join them. We show that this allows a coalitional move supported by a lower value of β than in the previous case and compute the predicted final split for all values of α.

Let $V = \{1, 2, ..., 34\}$ be the set of 34 members of Zachary's Karate Club. Let E be the set of relationships between the members and let $\Gamma = (V, E)$ be the associated graph. Let strategy A (resp. B) refer to joining the subclub led by member 1 (resp. 34). Before the split, all members of V play strategy B. Let S be the subset of V consisting of the members that were observed to play strategy A after the split. According to Zachary (1977),

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 17, 18, 20, 22\}.$$

We first consider the move by S from B to A as a single switch by a coalition comprising every member of S. The area of S is $d(S, S) = 33$. The perimeter of S is $d(S, V \setminus S) = 10$. At the initial strategy profile, all players play B, so there is no contaged perimeter and thus the perimeter equals the uncontaged perimeter. Consequently, the switch by S strictly increases the sum of payoffs of its players if and only if $\beta > 5/33$. Similarly, this switch is strictly potential increasing if and only if $\alpha > 10/33$. The least integrated member of S is member 3, so

$$Co(S) = \min_{i \in S} \frac{d(i,S)}{d(i)} = \frac{d(3,S)}{d(3)} = \frac{5}{10} = \frac{1}{2}. \tag{B.1}$$

By (B.1) and Remark 2, S is agency autonomous if and only if $\beta > 1$. That is, the coalitional move by S strictly increases the payoff of every member of S if and only if $\beta > 1$.

Next we consider the possibility that the move by S from B to A comprised a switch of strategy by some coalition $S_0 \subset S$, followed by switches by individual players in $S \setminus S_0$. Ideally the coalition S_0 would find it easier to switch from B to A than S would. That is, we would like S_0 to be agency autonomous for some $\beta < 1$. For this to be the case, we require $Co(S_0) > 1/2$. We further assume that $S_0 \supset V_0$, where V_0 is the set of all players $v \in V$ that satisfy the following property: every path in Γ between members v and 34 passes through member 1. Observe that $V_0 = \{1, 5, 6, 7, 11, 12, 17\}$.

Consider a dynamic process which begins with S_0 moving as a coalition from B to A such that the payoff of every player in S_0 strictly increases. This is followed by a sequence of individuals $v_i \in V$ for $i = 1, \ldots, n$ each moving from B to A and strictly increasing their respective payoffs. The process terminates when there does not exist any individual playing B who could strictly increase their payoff by moving to A. If the aforementioned conditions are satisfied, we say that the sequence *terminates* at $\hat{S}(S_0, \alpha) = S_0 \cup \{v_1, \ldots, v_n\}$. Observe that $\hat{S}(S_0, \alpha)$ is independent of β and the ordering of the sequence of v_i. In general, different choices of S_0 will lead to different $\hat{S}(S_0, \alpha)$. However, it turns out that all S_0 satisfying our conditions above lead to the same $\hat{S}(S_0, \alpha)$ for all values of α. Hence, without loss of generality we will assume that

$$S_0 := \{1, 2, 4, 5, 6, 7, 8, 11, 12, 14, 17, 18\},$$

for which $Co(S_0) = 5/9$. Hence, S_0 is agency autonomous if and only if $\beta > 4/5$. Furthermore, the move is potential increasing if and only if $\alpha > 15/22$. Note that although the initial switch by S_0 will not necessarily increase potential, every individual switch on the sequence $v_1, \ldots, v_n$ will increase potential. This is because, by definition of the potential function (2.2), an individual move is payoff increasing if and only if it is potential increasing.

Sequences obtained by this process are

$$\begin{array}{ll} S_0, 13, 20, 22, 3, 10 & \text{if } 0 < \alpha \leq 1/2 \\ S_0, 13, 20, 22, 3, 10, 9, 31 & \text{if } 1/2 < \alpha \leq 1 \\ S_0, 13, 20, 22, 3, 10, 9, 31, 29, 34, \ldots & \text{if } \alpha > 1 \end{array}$$

and so we obtain

$$\hat{S}(S_0, \alpha) = \begin{cases} S_0 \cup \{13, 20, 22, 3, 10\} & \text{if } 0 < \alpha \leq 1/2 \\ S_0 \cup \{13, 20, 22, 3, 10, 9, 31\} & \text{if } 1/2 < \alpha \leq 1 \\ V & \text{if } \alpha > 1 \end{cases}$$

as the subclub of V that splits from the original club. The prediction $\hat{S}(S_0, \alpha)$ for $0 < \alpha \leq 1/2$ differs from the observed subclub S only by the presence of member 10. Member 10 has only two neighbors in Γ, namely 3 and 34, hence a possible explanation for the absence of 10 in S could be that he had a stronger relationship with 34 that he had with 3. Referring to the weighted version of the network given on page 462 of Zachary (1977), this indeed does seem to be the case.

References

Ambrus, A., 2009. Theories of coalitional rationality. J. Econ. Theory 144, 676–695.
Angus, S.D., Newton, J., 2015. Emergence of shared intentionality is coupled to the advance of cumulative culture. PLoS Comput. Biol. 11, e1004587.
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.
Bacharach, M., 1999. Interactive team reasoning: a contribution to the theory of co-operation. Res. Econ. 53, 117–147.
Bernheim, B.D., Peleg, B., Whinston, M.D., 1987. Coalition-proof Nash equilibria I. Concepts. J. Econ. Theory 42, 1–12.
Blume, L.E., 1993. The statistical mechanics of strategic interaction. Games Econ. Behav. 5, 387–424.
Blume, L.E., 2003. How noise matters. Games Econ. Behav. 44, 251–271.
Bollobás, B., 2001. Random Graphs, 2nd edition. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, UK.
Candogan, O., Menache, I., Ozdaglar, A., Parrilo, P.A., 2011. Flows and decompositions of games: harmonic and potential games. Math. Oper. Res. 36, 474–503.
Dembo, A., Montanari, A., 2010. Ising models on locally tree-like graphs. Ann. Appl. Probab. 20, 565–592.
Gilbert, M., 1981. Game theory and convention. Synthese 46, 41–93.
Gillies, D.B., 1959. Solutions to general non-zero-sum games. Contrib. Theory Games 4, 47–85.
Goldberg, K., Goldman, A., Newman, M., 1968. The probability of an equilibrium point. J. Res. Natl. Bur. Stand. 72, 93–101.
Kajii, A., Morris, S.E., 1997. Refinements and higher order beliefs: a unified survey. University of Pennsylvania. Mimeo.
Lewis, D., 1969. Convention: A Philosophical Study. Harvard University Press, Cambridge, MA, USA.
Marden, J.R., Young, H.P., Pao, L.Y., 2014. Achieving Pareto optimality through distributed learning. SIAM J. Control Optim. 52, 2753–2770.
Monderer, D., Shapley, L.S., 1996. Potential games. Games Econ. Behav. 14, 124–143.
Moreno, D., Wooders, J., 1996. Coalition-proof equilibrium. Games Econ. Behav. 17, 80–112.
Morris, S., 2000. Contagion. Rev. Econ. Stud. 67, 57–78.
v.Neumann, J., 1928. Zur theorie der gesellschaftsspiele. Math. Ann. 100, 295–320.
Newton, J., 2012. Coalitional stochastic stability. Games Econ. Behav. 75, 842–854.

Newton, J., 2017. Shared intentions: the evolution of collaboration. Games Econ. Behav. 104, 517–534.
Newton, J., 2018. Evolutionary game theory: a renaissance. Games 9, 31.
Newton, J., 2019. Agency equilibrium. Games 10.
Newton, J., Angus, S.D., 2015. Coalitions, tipping points and the speed of evolution. J. Econ. Theory 157, 172–187.
Norman, T., 2009. Skew-symmetry under simultaneous revisions. Int. Game Theory Rev. 11, 471–478.
Peski, M., 2010. Generalized risk-dominance and asymmetric dynamics. J. Econ. Theory 145, 216–248.
Pradelski, B.S., Young, H.P., 2012. Learning efficient Nash equilibria in distributed systems. Games Econ. Behav. 75, 882–897.
Reich, B., 2016. The Diffusion of Innovations in Social Networks. Working Paper.
Rusch, H., 2019. The evolution of collaboration in symmetric 22-games with imperfect recognition of types. Games Econ. Behav. 114, 118–127.
Sawa, R., 2014. Coalitional stochastic stability in games, networks and markets. Games Econ. Behav. 88, 90–111.
Tomasello, M., 2014. A Natural History of Human Thinking. Harvard University Press, Cambridge, MA, USA.
Tomasello, M., Rakoczy, H., 2003. What makes human cognition unique? From individual to shared to collective intentionality. Mind Lang. 18, 121–147.
Ui, T., 2001. Robust equilibria of potential games. Econometrica 69, 1373–1380.
Young, H.P., 2011. The dynamics of social innovation. Proc. Natl. Acad. Sci. 108 (Suppl 4), 21285–21291.
Zachary, W.W., 1977. An information flow model for conflict and fission in small groups. J. Anthropol. Res. 33, 452–473.