



Decentralized learning from failure[☆]

Andreas Blume^a, April Mitchell Franco^{b,*}

^a*Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260, USA*

^b*Department of Economics, University of Iowa, Iowa City, IA 52242, USA*

Received 12 February 2002; final version received 27 January 2006

Available online 24 March 2006

Abstract

We study decentralized learning in organizations. Decentralization is captured through Crawford and Haller's [Learning how to cooperate: optimal play in repeated coordination games, *Econometrica* 58 (1990) 571–595] *attainability* constraints on strategies. We analyze a repeated game with imperfectly observable actions. A fixed subset of action profiles are *successes* and all others are *failures*. The location of successes is unknown. The game is played until either there is a success or the time horizon is reached. We partially characterize optimal attainable strategies in the infinite horizon game by showing that after any fixed time, agents will occasionally randomize while at the same time mixing probabilities cannot be uniformly bounded away from zero.

© 2006 Elsevier Inc. All rights reserved.

JEL classification: C72; C73; D23; D83

Keywords: Search; Decentralization; Symmetry; Attainability

[☆] We are grateful for comments made by Avinash Dixit, Paul Heidhues, Jim Jordan, Andy McLennan, Ariel Rubinstein, and seminar participants at Depaul University, Federal Reserve Bank of Minneapolis, Federal Reserve Bank of New York, 2002 Midwest Theory meetings, 2003 Murray S. Johnson Memorial Conference, Pennsylvania State University, Purdue University, 2001 Society for Economic Dynamics conference, New York University Stern School of Business, University of Arizona, University of Pittsburgh, Wissenschaftszentrum Berlin für Sozialforschung (WZB), 2004 Worldcongress of the Game Theory Society. We are indebted to the associate editor and referees of *JET* for their extraordinarily detailed comments and for pushing us toward more general results.

* Corresponding author. Fax: +319 335 1956.

E-mail addresses: ablume@pitt.edu (A. Blume), april-franco@uiowa.edu (A.M. Franco).

1. Introduction

We study decentralized learning in organizations. We require agents to learn by trial and error. They can only observe success or failure and not the actions taken by other agents. A failure means that the organization will want to explore a novel action combination. Decentralization means that agents change their actions independently. Therefore it may be unclear which members have to change their actions and, given the observational restrictions, which agents do change their actions. The learning activities of some agents may confound the learning of others.

Decentralization is captured through a symmetry constraint on agents' strategies, called *attainability*, that was introduced by Crawford and Haller [4] (CH in the sequel). An attainable strategy respects at any point in time whatever symmetries remain in the game. Among learning rules satisfying the symmetry constraints, we are interested in optimal rules. Formally, we analyze a repeated n -player game in which a fixed subset of action profiles are *successes* and all other action profiles are *failures*. The number of successes is known, but not their location. The game is played in periods $1, \dots, T$ until either there is a success or the time horizon, T , is reached.

Optimal attainable strategies are equilibrium strategies. Our main results show that with any optimal attainable strategy in the infinite horizon game, agents will never stop randomizing forever, while at the same time mixing probabilities cannot be uniformly bounded away from zero. Randomization makes it possible that players observe different histories and subsequently behave asymmetrically even if they use identical strategies. Randomization helps players getting "off the diagonal." The cost of randomization is that it may let players revisit action profiles by chance. An optimal strategy balances this cost with the desymmetrizing benefit from randomization. Thus randomization can be viewed as an investment in symmetry breaking.

We compare this, decentralized, solution to the centralized solution where the symmetry constraint is removed. In the infinite horizon game, the cost of decentralization, the difference between the unconstrained optimum and the attainable optimum, is always positive. It is (weakly) increasing with the division of labor in the organization, yet may fall with the size of the organization if the organization has a constant returns technology.

Crawford and Haller [4], who first proposed the approach that we are using in this paper, use it to analyze two-player repeated coordination games. It is extended in Kramarz [8] to n -player games, in Bhaskar [2] to games with conflict, and in Alpern and Reyniers [1] to dispersion games. Blume [3] uses analogous symmetry constraints to study the benefit of grammar-like structures in language. Rubinstein [15,14] uses an alternative approach for studying structure in language through properties of binary relations.

The paper is organized as follows. The next section describes search-for-success games. Section 3 introduces the attainability constraint on strategies that we use to capture decentralization. In Section 4 we solve for both optimal and equilibrium attainable strategies in a three-period two-action example. Our main results are in Section 5, where we characterize optimal and equilibrium attainable strategies in infinite horizon search-for-success games. In Section 6 we evaluate the expected payoff loss that results from conducting a decentralized rather than a coordinated search. Section 7 concludes.

2. The game

We consider a repeated n -player, m -action game. In the stage game, the players, $i \in I$ have identically many actions a_i and identical payoffs from each action combination $a = (a_1, \dots, a_n)$. Let A_i be player i 's set of actions in the stage game and note that $\#A_i = m$ for all i . All action

combinations are either *failures* or *successes*. Use X to denote the set of all possible assignments of successes to action profiles. Then, each player’s payoff $u_i(a, \theta)$ in the stage game depends on the action profile a and the random variable θ which takes values in X and determines which k profiles are successes. We assume that the number of success profiles k is the same for any realization of θ and satisfies $k < m^n - m$. The latter assumption ensures that a single player cannot guarantee a success by trying all of his actions. At a success, all players earn a payoff 1 and at a failure they all get 0. The number of successes is commonly known but not their location. Any assignment of the given number of successes across action combinations is equally likely. We refer to this game as “search for success.”

In the repeated game the random assignment of successes to action profiles is determined once-and-for-all before the first play of the game. The stage game is repeated in periods $t = 1, \dots, T$, until either a success is played or the time horizon T is reached.¹ We consider both finite and infinite T . Denote the repeated game with time horizon T by Γ^T . Players only observe their own actions and their own payoffs, not the actions of other players. We assume that players maximize the expected present discounted value of future payoffs with a common discount factor $\delta \in (0, 1)$.

Since players can only observe success or failure in addition to their own actions, and a success ends the game, their strategies can only depend on their private histories $h_i^t \in H_i^t$, where $H_i^t := A_i^t$ for $t \geq 1$, and we define H_i^0 as a singleton set containing the null history h_i^0 . Let $h^t := (h_1^t, \dots, h_n^t)$ be an entire history ending in period t and denote an infinite history by h . Define $H^t := \times_{i \in I} H_i^t$, and let H stand for the set of all infinite histories. A (behavior) strategy for player i is a sequence of functions

$$f_i^t : H_i^{t-1} \rightarrow \Delta(A_i),$$

where $\Delta(A_i)$ denotes the set of probability distributions over A_i . For any history $h_i^{t-1} \in H_i^{t-1}$ and $a_i \in A_i$, let $f_i^t(h_i^{t-1})(a_i)$ denote the probability that player i ’s strategy assigns to his action a_i after history h_i^{t-1} . Define $a_{i\tau}$ as the action taken by player i in period τ and $a_i^\tau(h_i^t)$ as the τ th component of the history h_i^t for $\tau \leq t$. Define the set of actions a_i taken by individual i in history h_i^t as $\mathcal{A}_i(h_i^t) := \{a_i \in A_i \mid \exists \tau \leq t \text{ with } a_i = a_i^\tau(h_i^t)\}$.

Any behavior strategy profile f induces a probability $\pi_t(f)$ of a success in period t , taking into account that a success ends the game. Note that it suffices to write this probability as $\pi_t(f^1, \dots, f^t)$, that is only as a function of behavior up to and including period t . We can then write each player’s payoff from profile f as

$$\Pi(f) = \sum_{t=1}^T \delta^{t-1} \pi_t(f^1, \dots, f^t).$$

For any two strategies f and g , we denote by $g(f, \tau)$ the strategy obtained by following f until period τ and g thereafter.

The fundamental problem in this game is to learn from failure, i.e. to avoid action combinations that have led to failures in the past. This is difficult when decision making is decentralized, communication is limited and agents lack other means of coordinating their search. In the next section, we formally model the constraints imposed on agents by decentralized decision making.

¹ For most of our results there is a version that holds for more standard repeated games that continue until the time horizon is reached regardless of the number of successes. However, our approach of ending the game as soon as there is a success considerably sharpens the statement of results.

3. Attainable strategies

Following Crawford and Haller [4], we use symmetry to capture the strategic uncertainty that players are facing in the absence of communication or of an alternative coordination mechanism. CH model players as having common knowledge of the game, while lacking a common description of the game. Players do not have a common language that would label either players' actions or their position in the game.

Like CH we focus on *attainable strategies*, i.e. strategies that are invariant to the players' private descriptions of the game. Up to a bijection between their action spaces, A_i and $A_{i'}$, any two players i and i' will use identical strategies, and, each player i 's strategy is only defined up to a permutation of A_i . Note that any bijection $\rho : A_i \rightarrow A_{i'}$ between two players' action spaces naturally induces a bijection between their spaces of histories H_i^t and $H_{i'}^t$ via $\rho(a_{i1}, \dots, a_{it}) = (\rho(a_{i1}), \dots, \rho(a_{it}))$. Then, the two conditions of *player symmetry* and *action symmetry* amount to

1. $\forall i, i' \in I, \exists$ a bijection $\hat{\rho} : A_i \rightarrow A_{i'}$ such that $f_i^t(h_i^{t-1})(a_i) = f_{i'}^t(\hat{\rho}(h_i^{t-1}))(\hat{\rho}(a_i))$, $\forall a_i \in A_i$, and
2. $f_i^t(h_i^{t-1})(a_i) = f_i^t(\tilde{\rho}(h_i^{t-1}))(\tilde{\rho}(a_i))$, $\forall a_i \in A_i$, and for all bijections $\tilde{\rho} : A_i \rightarrow A_i$.

We call strategy profiles f that satisfy both of these restrictions *attainable*. We say that an individual strategy f_i is attainable if it satisfies condition (2), i.e. action symmetry.

The second part of the definition of attainability implies that if a player puts any weight on an action that he has not used before, he must put equal weight on all such actions. Thus, in the first period, each player must put equal weight on each of his actions. In the second period, players can repeat their first-period actions, they can randomize equally over the new actions, or, more generally, they can play any convex combination of these two extremes. The first part of the definition of attainability implies that each player plays the same such convex combination; i.e., their second-period randomizations over actions are essentially the same. Suppose that in the second period the players put some weight on old and some weight on new actions. Then, in the third period, the players randomly divide into two groups: those whose first two actions were the same, and those whose first two actions differed. In the first group, each player's play in the third period must be a (possibly degenerate) mixture of the old action and new actions, with equal weight attached to each new action. In the second group, each player's play in the third period must be a (possibly degenerate) mixture of his first action, his second action and new actions, with equal weight attached to new actions. The mixtures within each group must be the same. So, for example, if some player in the second group puts weight $\frac{1}{2}$ on his first action, every player in this group must put weight $\frac{1}{2}$ on their respective first actions. But the mixtures can differ between the two groups. In general, after any history, an attainable strategy must specify a mixture of the actions that occurred in the player's history (with each action identified by when it occurred) and new actions, with the constraint that the weight on each new action must be equal. If two players' histories are the same (up to the names given to their actions by when the action occurred), then they must play the same such mixture.

The two parts of the definition of attainability are the same as CH's requirement that "players whose positions are undistinguished play identical strategies" and that "any two of a player's undistinguished actions enter his strategy symmetrically." Symmetries can only be broken over time, in the case of actions because they become distinguished by the period in which they are first taken, and in the case of players because, in our setting, they may observe different private histories.

Strategy profiles that maximize players' payoffs among the set of attainable strategy profiles will be called *optimal attainable strategy profiles*. The corresponding individual strategies will be referred to as optimal attainable strategies. Attainable strategy profiles that are Nash equilibria are referred to as *equilibrium attainable strategy profiles*. Note that when we check for equilibrium we impose *no* restrictions on deviations; the symmetry restrictions apply only on the solution path, whether it is optimal or equilibrium. Since players have to use identical strategies we will often limit our discussion to individual strategies rather than entire profiles. We denote the set of attainable behavior strategies of player i in a T -period search-for-success game by \mathcal{F}_i^T , and the set of attainable strategy profiles by \mathcal{F}^T . A private history h_i^t is *on the path* of an attainable strategy profile f if h_i^t has positive probability under f .

4. A three-period two-action example

In this section, we study the three-period game ($T = 3$) with two actions per player ($m = 2$), n players, $k < 2^n - 2$ success profiles and common discount factor $\delta \in (0, 1)$. The example illustrates the attainability constraint and builds intuition for our main results on infinite-horizon games.

With two actions, attainability amounts to players using both actions with equal probability in the first period and switching actions with identical probabilities following identical switching histories. Therefore, each player's strategy in the three-period game can be summarized by the triple (p, q_0, q_1) , where p is the probability of switching in period two, q_0 is the probability of switching in period three conditional on no switch in the previous period, and q_1 is the probability of switching in period three conditional on a switch in the previous period.

Before proceeding, we consider the case of a two-period game. Then, only the second-period switching probability p matters. By attainability, p must be the same for all players. If $p < 1$, there is positive probability, $(1 - p)^n$, that the action profile visited in the first period is revisited. Choosing $p = 1$ guarantees that a new action profile will be selected in the second period. Therefore, $p = 1$ is uniquely optimal in the two-period game.

In contrast, in the three-period game, $p = 1$ cannot be optimal. To see this, suppose that all players switch with probability one in the second period. Then all players have identical histories at the beginning of the third period and thus use identical switching probabilities in the third period. Therefore, there is positive probability that all players make the same switching decision in the third period and thus revisit a profile they have examined before. Intuitively, in this case they "stay on the diagonal." Suppose now that one of the players deviates by not switching in period two and then switches with probability one in the last period. In that case players will "get off the diagonal" in period two and a new profile is examined in every period. If all players use the original strategy with probability $1 - \varepsilon$ and adopt the deviation with probability ε , the resulting hybrid profile is attainable, and the probability that all players use the deviation, ε^n , is at least an order of magnitude smaller than the probability that a strict subset of players adopts the deviation. As a result, for sufficiently small ε , the hybrid profile would yield a strictly higher payoff than switching with probability one in period two.

Evidently, $p = 0$ cannot be optimal either, because then at most two profiles would be examined in total, one in period one and one in period three, while with discounting it would be better to examine these two profiles in the first two periods, i.e. $p = 1$. We conclude that an optimal value of p in the three-period game must satisfy $p \in (0, 1)$. The optimal p balances the fundamental tradeoff between increasing the probability that a new action profile is examined in period two and making it more likely that players start the third period off the diagonal. Before determining

the optimal second-period switching probability, we need to consider the third-period switching probabilities.

Conditional on not having switched in period two, it is optimal to switch with probability one in period three, i.e. $q_0 = 1$, because this guarantees that a new profile will be examined. Since there is positive probability that no player switches in period two, $q_0 = 1$ is uniquely optimal. Conditional on having switched in period two, there are two possibilities: all other players switched as well, or at least one other player did not switch in period two. In the latter case, the player who did not switch in period two will switch with probability one in the last period, because $q_0 = 1$, and therefore in this event the choice of q_1 does not affect the expected payoff. In the other case, where all players switched in period two, the probability of getting off the diagonal in period three equals $1 - q_1^n - (1 - q_1)^n$, which is uniquely maximized by setting $q_1 = \frac{1}{2}$. Since both cases have positive probability, the unique optimal choice for q_1 is $q_1 = \frac{1}{2}$.

We can now return to determining the optimal second-period switching probability p . Conditional on no success in period one, the ideal outcome in period two is for a strict subset of players to switch, because then it is guaranteed that a new profile is visited in every period. The expected payoff from this ideal outcome, starting in period two, equals $\frac{k}{2^{n-1}} + \delta \frac{2^n - 1 - k}{2^{n-1}} \frac{k}{2^{n-2}}$. An ideal outcome is not realized if one of two “errors” occurs: no one switches in period two, a “no-switch error;” or all players switch in period two, an “all-switch error.” The expected payoff from a no-switch error equals $\delta \frac{k}{2^{n-1}}$, using $q_0 = 1$. The expected payoff from an all-switch error equals $\frac{k}{2^{n-1}} + \delta \frac{2^n - 1 - k}{2^{n-1}} \frac{k}{2^{n-2}} (1 - (\frac{1}{2})^{n-1})$, using $q_1 = \frac{1}{2}$. The goal in choosing p is to minimize the expected payoff loss relative to the ideal outcome, which can now be calculated to be proportional to

$$\left((1 - \delta) + \delta \frac{2^n - 1 - k}{2^n - 2} \right) (1 - p)^n + \delta \frac{2^n - 1 - k}{2^n - 2} \left(\frac{1}{2} \right)^{n-1} p^n, \tag{1}$$

where the first term derives from the no-switch error and the second term from the all-switch error. There is a unique value of p that minimizes this expression. We summarize our observations in the following proposition:

Proposition 1. *In the three-period repeated search-for-success game with two actions per player, common discount factor δ , n players, and k success profiles, there exists a unique optimal attainable strategy. It is given by $(p, q_0, q_1) = (p^*(\delta, n, k), 1, \frac{1}{2})$, where*

$$p^*(\delta, n, k) = \frac{2}{\left(\frac{1}{\left(\frac{1-\delta}{\delta} \right) \left(\frac{2^n - 2}{2^n - 1 - k} \right) + 1} \right)^{\frac{1}{n-1}} + 2}.$$

The next question we will address is whether the unique optimal attainable strategy is an equilibrium strategy and, if so, whether there are other attainable equilibria. If the optimal attainable strategy were not an equilibrium, there would be a profitable deviation. Since players have identical payoffs, this deviation would benefit all players. All players using the original strategy with probability $1 - \varepsilon$ and the deviation with probability ε is an attainable strategy. The probability of exactly one player deviating is of higher order of magnitude than the probability of more players deviating. Therefore, for small ε the hybrid profile would lead to a higher expected payoff for all players, contradicting the optimality of the strategy we

found in Proposition 1. We conclude that the optimal attainable strategy of Proposition 1 is an equilibrium strategy.

Next, we turn to the question of uniqueness in the class of attainable equilibria. Consider the value of p in any attainable equilibrium. Suppose that $p = 1$. Then player i can guarantee that a novel action profile is examined in every period by defecting to the strategy $(0, 1, q_1)$. Suppose instead that $p = 0$. Then q_0 cannot be less than 1. Otherwise player i could deviate to $(0, 1, q_1)$ and thereby increase his payoff. However, if $p = 0$ and $q_0 = 1$, then player i would be better off by deviating to $(1, 1, q_1)$, which would guarantee that a novel profile is examined in every period. Hence, in any attainable equilibrium strategy, $p \in (0, 1)$.

Since $p < 1$, there is positive probability that no one switches in period two. It is always a best response for player i to switch in period three if he did not switch in period two, and if no one else switches in either period, it is a strict best response. Therefore q_0 must equal one in any attainable equilibrium strategy.

Since $p > 0$, the choice of q_1 matters with positive probability, and since $q_0 = 1$, it matters only when all players switch in period two. In that event it is uniquely optimal for player i to switch (not to switch) in period three if $q_1^{n-1} < (>)(1 - q_1)^{n-1}$, and he is indifferent otherwise. Hence, since attainability requires q_1 to be the same for all players, we must have $q_1 = \frac{1}{2}$.

Could there be a p other than $p^*(\delta, n, k)$ that is part of an attainable equilibrium? To answer this question, note that for player i the decision of whether to switch in period two is only relevant if either all or none of the other players switch. In all other cases, since $q_0 = 1$, it is guaranteed that a new profile is chosen in each period, regardless of i 's choice. If all other players switched with probability one, it would be optimal for player i to switch with probability zero, and vice versa. Therefore, there is a unique switching probability that makes player i indifferent between switching and not switching, which has to be $p^*(\delta, n, k)$.

We summarize our observations in the following proposition.

Proposition 2. *In the three-period repeated search-for-success game with two actions per player, common discount factor δ , n players, and k success profiles, there exists a unique attainable equilibrium strategy. It coincides with the optimal attainable strategy.*

Note that in the attainable Nash equilibrium, agents condition their behavior on their own past actions, which are not publicly observable. Since the attainable equilibrium is unique, it follows that there is no public attainable Nash equilibrium in our game.²

Simple inspection of $p^*(\delta, n, k)$ yields a number of interesting comparative statics predictions about how agents' responses to failure in an organization vary with the fundamentals that characterize the organization: first, the second-period switching probability $p^*(\delta, n, k)$ is strictly

² There has been some recent work on the effects of restricting the analysis of games with imperfect monitoring to public equilibria. Fudenberg et al. [6] obtain a Folk Theorem in perfect public strategies for games with imperfect public monitoring as long as the public signal permits statistical detection of individual deviations. In contrast, Radner et al. [13] show that perfect public equilibrium payoffs are uniformly bounded away (in the discount factor) from the efficient frontier in repeated partnership games with discounting, while Radner [12] shows that efficiency can be attained in repeated partnership games without discounting. Recently, Obara [11] has shown that the equilibrium payoffs obtainable with perfect public strategies in the Radner–Myerson–Maskin example can be improved upon by permitting players to condition their behavior on their own past actions, which are not publicly observable. Thus, Obara shows that the restriction to public strategies may constrain efficiency, while our example shows that this restriction in conjunction with the attainability requirement can rule out existence.

decreasing in δ . Intuitively, raising the discount factor shifts the concern from avoiding the no-switch loss to avoiding the all-switch loss. Agents become willing to increase the probability of a no-switch loss in order to increase the probability of getting off the diagonal. Second, as $\delta \rightarrow 1$, the second-period switching probability converges to $\frac{2}{3}$. To understand this note that there is with certainty a reduction in the number of profiles examined if no one switches in period two, whereas if all agents switch such a reduction occurs only with probability $\left(\frac{1}{2}\right)^{n-1}$. This means that agents will be more concerned with the immediate no-switch loss than the more remote all-switch loss. As can be seen from Eq. (1), when $\delta = 1$, minimizing the expected loss amounts to balancing $(1-p)^n$ and $2\left(\frac{1}{2}p\right)^n$. Thus the probability $(1-p)$ gets twice the weight of the probability p ; the minimum equates $(1-p)$ and $\frac{1}{2}p$. Third, as $\delta \rightarrow 0$, the second-period switching probability converges to 1. Very impatient players have no interest in the desymmetrizing effects of randomization, which only result in increased payoffs in the future. They seek immediate success. Fourth, the second-period switching probability is strictly decreasing in the number of agents, n , in the organization. An inspection of the expression in Eq. (1) reveals that increasing n magnifies the effect of small variations in p on the ratio of likelihoods of a no-switch versus and all-switch loss. This diminishes the relative importance of discounting and accordingly increases the concern with the all-switch loss, which leads to a reduction in p . This effect of increasing organization size on agents' responses to failure has the same direction even if the a priori success probability, $\frac{k}{2n}$, is kept constant. Fifth, the second-period switching probability is strictly increasing in k . Agents in the organization respond faster to failure as the success probability is increased. As k increases, it becomes less likely that period three is reached conditional on someone switching in period two. The increased probability of an immediate success diminishes the concern about the all-switch loss. Finally, the second-period switching probability converges to $\frac{2}{3}$ as $n \rightarrow \infty$. As we noted above, increasing the number of agents diminishes the role of discounting. The effect of letting $n \rightarrow \infty$ is the same as that of $\delta \rightarrow 1$.

5. Optimal attainable strategies

In this section we consider a general class of search-for-success games. As a preliminary step, we establish existence of optimal attainable strategies for games with any time horizon and a convenient link between optimal and equilibrium attainable strategies. We then proceed to the main objective of this paper, a partial characterization of optimal attainable strategies in infinite-horizon search-for-success games.

5.1. Existence of optimal attainable strategies and relation to equilibrium

Unlike in the previous section, with a long or infinite time horizon it is prohibitively difficult to explicitly exhibit optimal attainable strategies. Therefore, we choose to characterize such strategies instead. Such a characterization is only meaningful if we have a general existence result.

Proposition 3. *An optimal attainable strategy exists in any search-for-success game.*

When players have identical payoffs, as in search-for-success games, it is intuitive that optimal attainable strategy combinations are Nash equilibria. The following result confirms this intuition.

Proposition 4. *Any optimal attainable strategy in a search-for-success game is a Nash equilibrium strategy.*³

This result will play an important role in our characterization of optimal attainable strategies. To rule out a class of candidate optimal attainable strategies, by Proposition 4 it suffices to show that it does not contain a Nash equilibrium. The strategy of the proof is (similar to the argument in the example in Section 4) to show that if one player has a profitable deviation, it benefits all players to adopt that deviation with some small probability, which violates optimality.

Combining Propositions 3 and 4, we immediately obtain the following existence result for attainable equilibria.

Corollary 1. *An optimal attainable Nash equilibrium strategy exists in any search-for-success game.*

5.2. A (partial) characterization of optimal attainable strategies

We next provide a partial characterization of optimal attainable strategies in infinite horizon games. The first set of results in this section establishes the importance of randomization in optimal attainable strategies: we show that randomization never ceases (Proposition 5), that impatient players begin randomizing before they run out of new actions (Proposition 6) and that, similar to uniform mixing, an optimal attainable strategy finds a success with probability one (Proposition 7). The second set of results establishes the importance of near-deterministic behavior in optimal attainable strategies: we show that after no time can mixing probabilities be uniformly bounded away from zero (Proposition 8) and that with an increasing number of actions the probability of repeating an action during the initial periods converges to zero (Proposition 9).

The benefit of mixing is that it helps in desymmetrizing behavior, the cost that it may lead to repetitions by chance. A strategy whose mixing probabilities are bounded away from zero has the advantage that with an infinite time horizon it will find a success profile with probability one and thus for large δ will be nearly optimal. With such a strategy, however, there are sample paths along which the same action is taken repeatedly, without success, for a long time, making it increasingly unlikely for that action to be part of a success profile.

We begin by showing that any optimal attainable strategy requires nontrivial randomization. If players do not randomize, beyond the randomization that is implicit in the attainability requirement, they effectively use identical actions in each period. As a consequence, they will sample at most m profiles. More generally, if players cease randomizing after some time t , there is positive probability that, up to renaming, they all observed the same history until that time, in which case they will behave identically from time t on and once again only sample m profiles. In that case, eventually the players should become convinced that they are stuck in a cycle and would prefer to deviate from their strategy.

The next result makes this intuition precise. We first define a *deterministic* attainable strategy as one which, after each history, puts either probability one or probability zero on each used action. This implies that if after some history a deterministic strategy puts positive probability on an unused action, then it puts positive probability only on unused actions (and, by the attainability

³ CH and McLennan [9] prove analogous results for repeated coordination games without payoff uncertainty and for finite common interest games, respectively.

requirement assigns equal probability to each of them). More generally, we say that the attainable strategy f_i is *deterministic from time t* on if $f_i^\tau(h_i^{\tau-1})(a_i) \in \{0, 1\}$, for all $a_i \in \mathcal{A}_i(h_i^{\tau-1})$, for all $h_i^{\tau-1}$ on the path of f_i , and for all $\tau \geq t$.

Proposition 5. *For any $t < \infty$, an optimal attainable strategy f_i is not deterministic from time t on.*

Proof. Suppose to the contrary that the strategy f_i is deterministic after time t . Then for any history $h^t \in H^t$ there exists a time $\hat{t}(h^t, f)$ such that all action profiles that will be visited by f following h^t have been visited by time $\hat{t}(h^t, f)$. Since the set of action profiles is finite, $\hat{t}(h^t, f)$ is less than infinity. Define \hat{t} as the maximum of these times over all histories in H^t . Since H^t is finite, \hat{t} is finite as well. Note that the attainable profile f reaches time \hat{t} with positive probability. This will be the case for example when agents take identical actions (up to renaming) in each period up to time t . The strategy f_i is guaranteed to fail to find a success profile after time \hat{t} . Therefore, the strategy \tilde{f}_i that randomizes uniformly over all actions after time \hat{t} and is otherwise identical to f_i is both attainable and yields a strictly higher payoff than f_i , which contradicts f_i being an optimal attainable strategy. \square

Proposition 5 shows that regardless of their time preference players will want to continue randomizing some of the time in order to avoid getting stuck in cycles. Somewhat surprisingly, impatient players may want to randomize even before there is the risk of entering a cycle. To see this, suppose players always switched to an unused action as long as such an action is available, i.e. during the first m periods. Then a player who deviates and does not switch during one of the first m periods obtains the same expected payoffs for the first m periods and in addition can guarantee that a new profile will be visited during the $m + 1$ st period. Since the original strategy cannot avoid revisiting an earlier profile in period $m + 1$ with a probability that is bounded away from zero, the deviation increases expected payoffs during the first $m + 1$ periods. For sufficiently small discount factors, this payoff gain outweighs possible losses in future periods. In that case the original strategy is not an equilibrium and therefore not optimal by Proposition 4. This is made precise in the following proposition. The attainable strategy f_i *revisits a prior action during the first t periods* if there exists a period $\tau \leq t$ and an action $a_i \in \mathcal{A}_i(h_i^{\tau-1})$ such that $f_i^\tau(h_i^{\tau-1})(a_i) > 0$.

Proposition 6. *There exists a $\bar{\delta} \in (0, 1)$ such that for all $\delta \in (0, \bar{\delta})$ the following holds: If f_i is an equilibrium (optimal) attainable strategy in the game with m actions, then it revisits a prior action during the first m periods.*

Proof. Suppose f_i is an equilibrium attainable strategy that does not revisit a prior action during the first m periods. Let all players other than i follow this strategy (recall that in an attainable profile all players use identical strategies). Consider a deviation by player i to any strategy \tilde{f}_i with

$$\tilde{f}_i^t(h_i^{t-1}) = a_{i1} \quad \forall t = 1, \dots, m$$

and

$$\tilde{f}_i^{m+1}(h_i^m) = a_{i2},$$

where we use the fact that there are no attainability constraints on deviations. This deviation ensures that a novel action profile will be visited in each of the first $m + 1$ periods rather than only

the first m periods. Since the strategies used by players $j \neq i$ do not revisit a prior action during the first m periods,

$$\pi_t(\tilde{f}_i, f_{-i}) = \pi_t(f_i, f_{-i}) \quad \forall t \leq m.$$

It will be convenient to let $\psi_t(f)$ denote the hazard rate induced by the profile f , i.e.

$$\psi_t(f) := \frac{\pi_t(f)}{1 - \sum_{\tau=1}^{t-1} \pi_\tau(f)}.$$

Note that if every player follows the profile f during the first m periods then, given the attainability constraint, the probability of success in period $m + 1$ is maximized by solving the program

$$\min_p \sum_{j=1}^m p_j^n \quad \text{s.t.} \quad \sum_{j=1}^m p_j = 1,$$

where p is the vector of probabilities p_j that are assigned to actions a_{ij} , $j = 1, \dots, m$. The solution is

$$p_j^* = \frac{1}{m}.$$

Hence,

$$\psi_{m+1}(f) \leq \left(1 - m \left(\frac{1}{m}\right)^n\right) \frac{k}{m^n - m}.$$

In contrast,

$$\psi_{m+1}(\tilde{f}_i, f_{-i}) = \frac{k}{m^n - m}.$$

Therefore,

$$\begin{aligned} \pi_{m+1}(\tilde{f}_i, f_{-i}) - \pi_{m+1}(f) &= \left(1 - \sum_{\tau=1}^m \pi_\tau(f)\right) \left[\psi_{m+1}(\tilde{f}_i, f_{-i}) - \psi_{m+1}(f)\right] \\ &\geq \left(1 - \sum_{\tau=1}^m \pi_\tau(f)\right) \left(\frac{1}{m}\right)^{n-1} \frac{k}{m^n - m} \\ &=: \alpha > 0. \end{aligned}$$

Even if we give f the best shot at being optimal by making the most favorable assumption about its continuation payoff after period $\hat{t} = m + 1$,

$$\begin{aligned} \Pi(\tilde{f}_i, f_{-i}) - \Pi(f) &\geq \delta^{\hat{t}} \alpha - \frac{\delta^{\hat{t}+1}}{1 - \delta} \\ &= \delta^{\hat{t}} \left(\alpha - \frac{\delta}{1 - \delta}\right). \end{aligned}$$

Hence, if we define $\bar{\delta} := \frac{\alpha}{1+\alpha}$, then \tilde{f}_i is a profitable deviation as long as $\delta < \bar{\delta}$. \square

According to Proposition 5, continued occasional randomization will be an essential part of any optimal attainable strategy. It aids symmetry breaking and ensures that there are no profiles

that are systematically avoided. Indeed, as our next result shows, an optimal attainable strategy will eventually find a success.

Proposition 7. *An optimal attainable strategy f_i finds a success with probability one.*

Proof. Let $\zeta > 0$ denote the probability that a success will be found in the one-shot game if all players randomize uniformly over all their actions. The payoff from an optimal attainable strategy must be strictly greater than ζ . Define $\beta(1, t)$ as the probability that the optimal attainable strategy f_i finds a success in the time interval from time 1 to time t . $\beta(1, t) + \delta^{t+1}$ is weakly greater than the payoff from an optimal attainable strategy for all t . Choose t and ε such that $\zeta - \delta^{t+1} > \varepsilon > 0$. Then, $\beta(1, t) > \varepsilon$. More generally, the probability $\beta(lt + 1, lt + t)$ that a success will be found in the t periods starting with period $lt + 1$ conditional on no success in the preceding periods has to satisfy $\beta(lt + 1, lt + t) > \varepsilon \forall l \geq 0$. Therefore for all $l \geq 0$, the probability of a success in the infinite horizon game is not less than $1 - (1 - \varepsilon)^l$, which proves our claim. \square

While randomization is essential, mixing probabilities cannot be uniformly bounded away from zero. This can easily be seen in the case of continued uniform randomization. There will be sample realizations in which a player chooses a given action in every period for some length of time. If that length of time is large enough, the posterior probability of that action being part of a success profile will be arbitrarily close to zero. At that point the player is better off putting probability zero on that action. We proceed with formalizing and generalizing this intuition.

We say that f_i is ε -mixed after time t if for all $\tau > t$ and $h_i^{\tau-1}$ on the path of f_i , $f_i^\tau(h_i^{\tau-1})(a_i) > \varepsilon$ for all $a_i \in A_i$. We argue next that under an ε -mixed profile, there is positive probability that a player uses the same action for long periods of time, and as a consequence places arbitrarily small probability on the possibility of that action being part of a success profile. We conclude that after such a history the player is better off not using that action in at least one period, which contradicts the assumption that his equilibrium strategy is ε -mixed.

Proposition 8. *For all $\varepsilon > 0$, $\delta \in (0, 1)$ there does not exist an equilibrium attainable strategy f_i and a time t such that f_i is ε -mixed after time t .*

Proof. In order to derive a contradiction, suppose that f_i is an equilibrium attainable strategy that is ε -mixed after time t . Then it is optimal for player i to take the same action, say \tilde{a}_i , in every period following time t . Since all other players use identical strategies, their strategies are also ε -mixed after time t . Therefore the posterior probability of \tilde{a}_i being part of a success profile converges to zero if player i continues to take that action. Consequently, the expected payoff from continuing to take action \tilde{a}_i converges to zero as well. In contrast, always taking one of the actions that maximizes the instantaneous payoff yields an expected payoff that is bounded away from zero, which contradicts the optimality of taking action \tilde{a}_i in every period following time t . \square

Combining Propositions 8 and 4, we obtain the following observation.

Corollary 2. *For all $\varepsilon > 0$, $\delta \in (0, 1)$ there does not exist an optimal attainable strategy f_i and a time t such that f_i is ε -mixed after time t .*

An additional consequence of Proposition 8 is that optimal attainable strategies must exhibit nondegenerate dependence on private histories. They cannot be public, i.e. condition only on

publicly available information, which is the current time period. To see this, let $\rho : A_i \rightarrow A_i$ be a bijection, with $\rho(a_{i1}) = a_{ij}$ for some $j \neq 1$. By attainability,

$$f_i^t(h_i^{t-1})(a_{i1}) = f_i^t(\rho(h_i^{t-1}))(a_{ij}).$$

If f is public, then

$$f_i^t(h_i^{t-1}) = f_i^t(\rho(h_i^{t-1})).$$

Combining the two equations and using the fact that the choice of a_{ij} was arbitrary, we have

$$f_i^t(h_i^{t-1})(a_{i1}) = f_i^t(h_i^{t-1})(a_{ij}) \quad \forall a_{ij} \in A_i.$$

This shows that any public attainable strategy must always mix uniformly. By Proposition 8 we know that uniform mixing in every period is not an equilibrium. Hence, by Proposition 2 we know that the unique public attainable strategy is not an optimal attainable strategy.

While the extremes of uniform mixing in every period and of deterministic switching in every period are never optimal, these strategies will be approximately optimal for equally extreme choices of some of the parameters of the game. Evidently, for any $\varepsilon > 0$ we can find $\underline{\delta} \in (0, 1)$ such that for all $\delta > \underline{\delta}$, the payoff from uniform mixing is at least ε -close to that from an optimal strategy. In contrast, for fixed δ , switching with probability one to a new action as long as such actions are available (“switching early”) becomes increasingly attractive as the number of available actions in the game increases. This is similar to the case of a finite time horizon T , where as long as $m > T$ any strategy that revisits an action taken before would be dominated by a strategy that switches to a new action in every period. With an infinite horizon, increasing the number of actions affects the payoffs from switching early vis-à-vis early randomization in two opposing ways: the costs from not achieving desymmetrization when switching early can be postponed into the more distant future; at the same time, the costs of revisiting actions taken before when randomizing early decline because there is a lower *ex ante* probability of these actions being success profiles. The former costs fall exponentially with m because of discounting; the latter fall polynomially, because the *ex ante* probability of a success is a polynomial function of m . Therefore, for large m switching early becomes increasingly attractive. The following proposition makes this intuition precise.

Proposition 9. *Let f^m be an optimal attainable strategy profile for the infinite horizon game with m actions. Consider any sequence $\{f^m\}_{m=1}^\infty$ and denote by ϕ_t^m the corresponding probability with which period t is reached and f_i^m revisits the set of prior actions in period t . Then, for any $t < \infty$,*

$$\lim_{m \rightarrow \infty} \phi_t^m = 0.$$

6. The cost of decentralization

The attainability constraint in our analysis makes explicit that agents’ decision making is decentralized, i.e. that they cannot rely on communication, prior history or another mechanism to coordinate their strategies. In this section we ask what cost this constraint imposes on agents. Specifically, we define the *cost of decentralization* as the difference between the optimal unconstrained equilibrium payoff and the optimal attainable payoff.

A fundamental fact about the cost of decentralization follows directly from the attainability constraint. For any time $t \leq T$, there is positive probability that players use identical actions in

every period until time t . For $t > m$ this means that players will revisit action profiles with positive probability. Therefore we have the following observation:

Proposition 10. *The cost of decentralization in the search-for-success game is positive if and only if the time horizon T is greater than the number of actions m .*

We next turn to the question of how the cost of decentralization changes with the number of members of the organization. First-order intuition would suggest that increasing n complicates coordination and thereby increases the cost of decentralization. Note, however, that increasing n , without making any other changes, also reduces the payoff in the one-shot game, which confounds the analysis. It appears more appropriate to keep the ex ante productivity of the organization $\gamma = \frac{k}{m^n}$ (the chance of having a success in the one-shot game) constant as we vary the number of members of the organization.

There are two basic ways of keeping γ constant as we raise n : we can increase the number of success profiles k , which corresponds to *constant returns to scale* in the one shot game, or we can reduce the number of actions m , which, as we will see, can be interpreted as *increasing the division of labor* in the organization. Consider reducing m while raising n first. For example, compare the game with $n = 2$ players and $m = 4$ actions per player with the game with $n' = 4$ players and $m' = 2$ actions per player. Observe that the number of action profiles is the same in both games. Note also that we can identify (partial) action profiles in the four-player game with actions in the two-player game as follows: call player i 's actions in the four-player game $\{b_i, c_i\}$, $i = 1, 2, 3, 4$. Then we can “derive” the two-player game by calling player j 's actions in the two-player game $\{B_j, C_j, D_j, E_j\}$ and making the identification $B_1 = (b_1, b_2)$, $C_1 = (b_1, c_2)$, $D_1 = (c_1, b_2)$, $E_1 = (c_1, c_2)$ and $B_2 = (b_3, b_4)$, $C_2 = (b_3, c_4)$, $D_2 = (c_3, b_4)$, $E_2 = (c_3, c_4)$. Thus player $i = 1$ in the two-player game can be thought of as controlling the actions of players $j = 1$ and 2 in the four-player game, while player $i = 2$ in the two-player game can be thought of as controlling the actions of players $j = 3$ and 4 in the four-player game. If a player in the two-player controls the actions of two players in the four-player game, there is less division of labor in two-player game.

More generally, we can double the number of players from any n to $n' = 2n$, and at the same time reduce the number of actions m in order to keep the ex ante productivity γ constant.⁴ Then we can identify each player in the n -player game with a pair (i, j) in the n' -player game. It is convenient to give this compound player the name (i, j) and to think of him as choosing action profiles (a_i, a_j) in the n' -player game. The compound player (i, j) 's strategy assigns probabilities to actions profiles as a function of histories of action profiles, as follows:

$$f_{(ij)}(h_i^{t-1}, h_j^{t-1})(a_i, a_j).$$

In contrast, the probability of that same pair of actions if chosen by two different players in the n' -player game equals

$$f_i(h_i^{t-1})(a_i) \times f_j(h_j^{t-1})(a_j).$$

The set of attainable mappings from history profiles (h_i^{t-1}, h_j^{t-1}) into action profiles is strictly smaller for the pair of players i and j than for the compound player (i, j) for three reasons: (1) for some histories, each of the disjoint players i and j have access to strictly less information than the

⁴ We consider the case where the initial values of m and n are such that no integer problems arise.

compound player (i, j) , (2) unlike the disjoint players i and j , the compound player can correlate his component actions, and (3) the attainability constraints are more stringent for the disjoint players than for the compound player. The set of attainable mappings from history profiles to action profiles corresponds to the strategic options that are available to players in the game under the decentralization constraint. This construction generalizes straightforwardly to the case where n' is any multiple of n and numbers of actions in both games satisfy the condition $m^n = (m')^{n'}$. We refer to any such an increase in n , as an increase in the division of labor. Therefore, we have the following result:

Proposition 11. *Increasing the division of labor (1) strictly reduces the attainable strategic options in the game, and (2) weakly reduces expected payoffs from optimal attainable strategies.⁵*

Next, we examine how the cost of decentralization varies with n when there are constant returns in the one-shot game. Unfortunately, it is more difficult to obtain general results here. Therefore we return to our three-period two-action example from Section 4. There we can calculate the payoff from the optimal attainable strategy explicitly as

$$\pi^D = \frac{k}{2^n} + \frac{2^n - k}{2^n} \delta \left(\frac{k}{2^n - 1} (1 - (1 - p)^n) \right) + \frac{2^n - k}{2^n} \frac{k}{2^n - 1} \delta^2 \left\{ (1 - p)^n + \frac{2^n - 1 - k}{2^n - 2} (1 - (1 - p)^n - 2^{1-n} p^n) \right\},$$

where the superscript D emphasizes that this is the payoff from decentralization and p is understood to be the optimal second-period switching probability derived in Section 4, which itself depends on the parameters δ , k and n .

In the centralized organization, members can communicate and thus can ensure that an action profile is never revisited. This implies that the expected payoff from centralization in the three-period two-action game equals

$$\pi^C = \frac{k}{2^n} + \delta \left(1 - \frac{k}{2^n} \right) \left(\frac{k}{2^n - 1} \right) + \delta^2 \left(1 - \frac{k}{2^n} \right) \left(1 - \frac{k}{2^n - 1} \right) \left(\frac{k}{2^n - 2} \right).$$

Therefore the costs of decentralization in the three-period two-action example equals

$$C(\delta, k, n) = \pi^c - \pi^d = \delta \left(\frac{2^n - k}{2^n} \right) \left(\frac{k}{2^n - 1} \right) \left\{ [(1 - \delta) (1 - p)^n] + \delta \left(\frac{2^n - 1 - k}{2^n - 2} \right) (2^{1-n} p^n - (1 - p)^n) \right\}.$$

Examining this cost function, we find that in the three-period repeated search-for-success game with two actions per player *the cost of decentralization*, $C(\delta, k, n)$ (1) is decreasing in n for large n if one keeps $\frac{k}{2^n}$ constant, and (2) converges to 0 as $n \rightarrow \infty$, with or without keeping $\frac{k}{2^n}$ constant. Unlike in Proposition 11, where raising n signifies an increase in the division of labor that makes it more difficult to avoid action profiles visited earlier, here raising n makes it easier to avoid repetition. The reason is that here raising n increases the number of action profiles. As the number

⁵ Jovanovic and Nyarko [7] also study a model in which increasing complexity slows learning. While considering multiple decision makers, their learning is centralized. Slower learning in their setting follows directly from increasing the dimensionality of the problem, not from coordination issues.

of action profiles grows, it becomes increasingly unlikely that a mixing strategy revisits an earlier profile in the three-period game.

In addition, we find that $C(\delta, k, n)$ (1) is strictly increasing in δ provided $\delta \in [0, \frac{1}{2})$ or $k = 1$, (2) converges to 0 as $\delta \rightarrow 0$, and (3) is strictly increasing in k for $(\frac{2^n - 1}{3}) > k$. Intuitively, (1) as the future becomes more important, the decentralization cost increases because the problem of revisiting action profiles that have been examined before becomes more acute. (2) If the future does not matter at all, the decentralization cost falls to zero because short-run payoffs are maximized by all agents switching to a new action. (3) With a moderate number of success profiles, the first-order effect of increasing the number of success profiles is to increase the payoffs with and without communication in the same proportion, which raises the difference. With a large number of success profiles this effect disappears because the probability of an early success becomes high, which makes the concern about revisiting prior profiles irrelevant.

7. Conclusion

We have studied decentralized search for an optimal action profile in common-interest games where players cannot use communication to coordinate strategies and cannot observe the actions of other players. We find that players invest in breaking symmetry through randomization. Players occasionally randomize after any time horizon, while at the same time mixing probabilities cannot be uniformly bounded away from zero.

Future work within the present symmetric game framework might explore generalizations of the payoff structure, e.g. games in which some successes are more valuable than others and games where in addition these valuations differ across players. More generally, it would be desirable to study decentralized search for an optimal profile in asymmetric games, where for example different players have different numbers of actions. This would necessitate finding a generalization or substitute for the attainability constraint that we presently use to capture decentralization; one possibility would be to introduce private information about which action profiles are more likely to lead to a success.

Appendix. A

A.1. Proof of Proposition 3

Proof. At a given information set, a behavior strategy specifies a probability distribution with finite support. The space of such probability distributions is compact. The attainability constraints at that information set are linear. Therefore, at any given information set, the set of probability distributions induced by an attainable behavior strategy is a closed subset of a compact set, and therefore compact.

The space \mathcal{F}^T of attainable behavior strategy profiles is isomorphic to the space \mathcal{F} of individual attainable behavior strategies which itself is a product of the attainable probability distributions at each information set. Therefore by Tychonoff's Theorem \mathcal{F}^T is compact in the product topology.⁶

In the finite horizon case the payoff function $\Pi(\cdot)$ is clearly continuous relative to the product topology. Hence, in this case, finding an optimal attainable strategy amounts to maximizing a continuous function over a compact set. A solution exists by Weierstrass' Theorem.

⁶ For this and later references to topology see e.g. Munkres [10].

In order to show existence in the infinite horizon game, we use the fact that with bounded payoffs in the stage game and constant common discount factor $\delta < 1$, the search-for-success game is *continuous at infinity* (see [5]), i.e. behavior in the far distant future has a vanishing effect on payoffs. Formally, if $V^i(x, h)$ is the payoff of player i in the infinite horizon game as a function of the realization x of the random variable θ and the history h and $h(\tau)$ denotes the truncation of the infinite history h after period τ , then the game is continuous at infinity if

$$\sup_{i \in I, x \in X, h, h' \in H, h(\tau) = h'(\tau)} |V^i(x, h) - V^i(x, h')| \rightarrow 0 \quad \text{as } \tau \rightarrow \infty.$$

As a consequence of continuity at infinity we have the following property: for any $\varepsilon > 0$, there exists τ_0 such that

$$\sup_{f, g} |\Pi(f) - \Pi(g(f, \tau))| < \varepsilon \quad \forall \tau > \tau_0.$$

For the infinite horizon case, note that since \mathcal{F}^∞ is a countable product, \mathcal{F}^∞ endowed with the product topology is metrizable. Recall that in a metrizable space compactness implies sequential compactness.

Let f_T^* be an optimal attainable strategy in Γ^T for finite T . Let \tilde{f}_T^* be the extension of f_T^* to Γ^∞ that is obtained by prescribing uniform randomization at every information set after time T . Since \mathcal{F}^∞ is sequentially compact, the sequence $\{\tilde{f}_T^*\}_{T=1}^\infty$ has a convergent subsequence. Denote this sequence (after reindexing) by $\{\tilde{f}_T^*\}_{T=1}^\infty$ as well and its limit by \tilde{f} . Note that \tilde{f} is attainable.

In order to obtain a contradiction, suppose there exists an attainable profile \hat{f} and an $\varepsilon > 0$ such that $\Pi(\hat{f}) - \Pi(\tilde{f}) > \varepsilon$. Let \hat{f}_T and \tilde{f}_T denote the strategies obtained from \hat{f} and \tilde{f} by truncating after T periods and prescribing uniform randomization thereafter. Since the payoff function is continuous at infinity, there exists a \underline{T} such that for all $T > \underline{T}$, we have

$$\Pi(\hat{f}_T) - \Pi(\tilde{f}_T) > \frac{\varepsilon}{2}.$$

$\Pi(\tilde{f}_n^*)$ is a bounded increasing sequence and therefore converges. Denote the limit by Π^* . Let $\tilde{f}_{n,T}^*$ denote the strategy obtained from \tilde{f}_n^* by truncating after period T and prescribing uniform randomization thereafter. Because convergence in the product topology implies pointwise convergence, for any T , $\tilde{f}_{n,T}^*$ converges to \tilde{f}_T . Since Π is continuous in the arguments referring to the first T periods, $\Pi(\tilde{f}_{n,T}^*)$ converges to $\Pi(\tilde{f}_T)$. Combining this with continuity at infinity of the payoff function implies that $\Pi(\tilde{f}) = \Pi^*$. Hence, there exists a T such that $|\Pi(\tilde{f}_T) - \Pi(\tilde{f}_T^*)| < \frac{\varepsilon}{4}$.

$$\Rightarrow \Pi(\hat{f}_T) - \Pi(\tilde{f}_T) - (\Pi(\tilde{f}_T^*) - \Pi(\tilde{f}_T)) > \frac{\varepsilon}{4}$$

and thus

$$\Pi(\hat{f}_T) - \Pi(\tilde{f}_T^*) > 0,$$

which contradicts the optimality of f_T^* in the game with time horizon T . \square

A.2. Proof of Proposition 4

Proof. The proof has two major steps. In the first step it is shown that if there is a profitable deviation, then there is a profitable attainable deviation. f^* is an optimal attainable strategy

profile if it solves $\max_{f \in \mathcal{F}^T} \Pi(f)$. If f^* is not a Nash equilibrium, then there exists a pure strategy \hat{f}_i for player i such that

$$\Pi(\hat{f}_i, f_{-i}^*) > \Pi(f^*).$$

This pure deviation generates a sequence of actions $\hat{a}_i^T = \{\hat{a}_{i\tau}\}_{\tau=0}^T$. Call this action sequence a “profitable deviation sequence.” Denote the initial segment until period $t < T$ by \hat{a}_i^t . Since player one’s actions a priori differ only in their names, for any permutation ρ of player one’s actions, the sequence $\rho(\hat{a}_i^T) = \{\rho(\hat{a}_{i\tau})\}_{\tau=0}^T$ also constitutes a profitable deviation sequence. Hence, any strategy that is certain to generate a sequence $\rho(\hat{a}_i^T)$ for some ρ is a profitable deviation. The following strategy, g_i , does that and is constructed to be attainable (i.e., it satisfies condition (2) of the definition of attainability):

$$g_i^t(h_i^{t-1})(a_i) = \begin{cases} 1 & \text{if } \exists \rho \text{ such that } \rho(\hat{a}_i^t) = (h_i^{t-1}, a_i) \\ & \text{and } a_i \in \mathcal{A}_i(h_i^{t-1}), \\ 0 & \text{if } \exists \rho \text{ such that } \rho(\hat{a}_i^{t-1}) = h_i^{t-1} \text{ and } \nexists \rho \text{ such that} \\ & \rho(\hat{a}_i^t) = (h_i^{t-1}, a_i), \\ \frac{1}{m - \#\mathcal{A}_i(h_i^{t-1})} & \text{if } \exists \rho \text{ such that } \rho(\hat{a}_i^t) = (h_i^{t-1}, a_i) \\ & \text{and } a_i \notin \mathcal{A}_i(h_i^{t-1}), \\ \frac{1}{m} & \text{if } \nexists \rho \text{ such that } \rho(\hat{a}_i^{t-1}) = h_i^{t-1}. \end{cases}$$

The first three lines of the specification of g_i deal with the case where the history thus far is a permutation of an initial segment of the original profitable deviation sequence. The first line says that if the original deviation sequence requires the repetition of an earlier action, then g_i assigns probability one to the permutation of the action that continues the permuted sequence. The second line complements the first line by requiring that under the same conditions the remaining actions are used with probability zero. The third line says that if the original deviation sequence next uses an action that has not been used before, then g_i prescribes uniform randomization over all actions that have not been used in the realized history. The fourth line ensures that the strategy is well defined by specifying uniform randomization after unreached histories. Hence, without loss of generality, we may take any profitable deviation \hat{f}_i of player i to be attainable.

In the second step of the proof it is shown that if there is a profitable deviation, then all players benefit from adopting it with small probability, which violates optimality. For each player i , consider a strategy \tilde{f}_i in which player i plays f_i^* with probability $1 - \varepsilon$ and \hat{f}_i with probability ε . We can construct a behavior strategy \tilde{g}_i that is outcome equivalent to \tilde{f}_i . For this purpose, denote by $\text{Prob}(h_i^t | f_i)$ the probability of the history h_i^t under the strategy f_i , when we choose a placement of success profiles and a profile of pure strategies of players other than i such that given i ’s sequence of actions in h_i^t there is no success before time t . Observe that this probability is well-defined, i.e. independent of the details of the choice of success placements and other players’ strategies. Say that a strategy f_i for player i reaches history h_i^t if $\text{Prob}(h_i^t | f_i) > 0$. Then \tilde{g}_i can be defined as follows:

1. For any history h_i^{t-1} that is reached with positive probability by either \hat{f}_i or f_i^* , let

$$\begin{aligned} & \tilde{g}_i^t(h_i^{t-1})(a) \\ & := \frac{\hat{f}_i^t(h_i^{t-1})(a)\text{Prob}(h_i^{t-1} | \hat{f}_i)\varepsilon + f_i^{*t}(h_i^{t-1})(a)\text{Prob}(h_i^{t-1} | f_i^*)(1 - \varepsilon)}{\text{Prob}(h_i^{t-1} | \hat{f}_i)\varepsilon + \text{Prob}(h_i^{t-1} | f_i^*)(1 - \varepsilon)}. \end{aligned}$$

2. For any history h_i^{t-1} that is reached with zero probability by both \hat{f}_i and f_i^* , let

$$\tilde{g}_i^t(h_i^{t-1})(a) := \frac{1}{m} \quad \forall a \in A_i.$$

The first part of the definition of \tilde{g}_i ensures that at the information set h_i^t the mixing probabilities are the same as the ones induced by \tilde{f}_i conditional on reaching that information set. The remainder of the definition is needed to guarantee attainability at unreached information sets.

The behavior strategy \tilde{g}_i inherits attainability from \hat{f}_i and f_i^* because (1) for any attainable strategy f_i , $\text{Prob}(h_i^{t-1} | f_i)$ is invariant under renaming of actions and (2) by attainability of \hat{f}_i and f_i^* the probabilities $\hat{f}_i^t(h_i^{t-1})(a)$ and $f_i^{*t}(h_i^{t-1})(a)$ are invariant under renaming of actions. Note that if each agent uses \tilde{g}_i , we get the same expected payoff as if each agent randomized once at the beginning of the game choosing the behavior strategy \hat{f}_i with probability ε and the behavior strategy f_i^* with probability $1 - \varepsilon$.

Denote the common payoff from m players using strategy \hat{f} and the remaining $n - m$ players using strategy f^* by $\Pi^{m,n}(\hat{f}; f^*)$. Then the common expected payoff from \tilde{f} equals

$$\begin{aligned} \Pi(\tilde{f}) &= \sum_{m=0}^n \binom{n}{m} \varepsilon^m (1 - \varepsilon)^{n-m} \Pi^{m,n}(\hat{f}; f^*) \\ &= (1 - \varepsilon)^n \Pi(f^*) + n\varepsilon(1 - \varepsilon)^{n-1} \Pi(\hat{f}_1, f_{-1}^*) \\ &\quad + \sum_{m=2}^n \binom{n}{m} \varepsilon^m (1 - \varepsilon)^{n-m} \Pi^{m,n}(\hat{f}; f^*). \end{aligned}$$

For small ε the third term on the right of the equation is at least a magnitude smaller than the second term. Hence,

$$\Pi(\tilde{f}) > \Pi(f^*)$$

for sufficiently small ε , which contradicts our assumption of f^* being an optimal attainable strategy. \square

A.3. Proof of Proposition 9

Proof. Suppose not. Then there exists a time t , an $\eta' > 0$ and a subsequence $\{f^{m_j}\}_{j=1}^\infty$ such that along the subsequence $\phi_t^{m_j} \geq \eta' > 0 \forall m_j$. After reindexing, we can write the subsequence itself as $\{f^m\}_{m=1}^\infty$. By condition (2) of the definition of attainability, player i 's strategy can be entirely described in terms of histories that label actions by the time at which they were taken, or equivalently by identifying histories that differ only in terms of the names of player i 's actions. Denote these reduced histories by λ_i^t . Since $\phi_t^m \geq \eta' > 0 \forall m$, for each m there exists a history $\lambda_i^{t,m}$ for player i whose probability is bounded away from zero and after which the probability of returning to the set of prior actions is bounded away from zero. This implies that the probability that each player i observes the history $\lambda_i^{t,m}$ and returns to the set of prior actions is bounded away from zero by some $\eta > 0$, because the probability of not finding a success in the first t periods is bounded away from zero and the probability of observing the action sequence $\lambda_i^{t,m}$ conditional on no success in the first t periods is bounded away from zero. Conditional on all players revisiting the set of prior actions and having experienced identical histories, the probability of them revisiting a prior action profile is bounded below by $\left(\frac{1}{t}\right)^n$, where this lower bound varies with t because the

maximum number of actions that players may have taken before period t increases in t . Combining these observations, in any equilibrium along the subsequence, the probability of revisiting a prior profile in period t is at least $\eta \left(\frac{1}{t}\right)^n$. Consider any sequence $\{g^m\}_{m=1}^\infty$ of attainable strategy profiles g^m that never revisit a prior action profile during the first m periods. Then,

$$\Pi(g^m) - \Pi(f^m) \geq \delta^t \eta \left(\frac{1}{t}\right)^n \left(\frac{1}{m}\right)^n - \delta^m \frac{1}{1 - \delta}.$$

For sufficiently large m , this payoff difference is strictly positive. This contradicts the assumed optimality of f^m for all m . \square

References

- [1] S. Alpern, D.J. Reyniers, Spatial dispersion as a dynamic coordination problem, *Theory Dec.* 53 (2002) 29–59.
- [2] V. Bhaskar, Egalitarianism and efficiency in repeated symmetric games, *Games Econ. Behav.* 32 (2000) 247–262.
- [3] A. Blume, Coordination and learning with a partial language, *J. Econ. Theory* 95 (2000) 1–36.
- [4] V. Crawford, H. Haller, Learning how to cooperate: optimal play in repeated coordination games, *Econometrica* 58 (1990) 571–595.
- [5] D. Fudenberg, D. Levine, Subgame-perfect equilibria of finite and infinite horizon games, *J. Econ. Theory* 31 (1983) 251–268.
- [6] D. Fudenberg, D. Levine, E. Maskin, The Folk theorem with imperfect public information, *Econometrica* 62 (1994) 997–1040.
- [7] B. Jovanovic, Y. Nyarko, A Bayesian learning model fitted to a variety of empirical learning curves, *Brookings Pap. Econ. Act.* 1 (1995) 247–299.
- [8] F. Kramarz, Dynamical focal points in N-Person coordination games, *Theory Dec.* 40 (1996) 277–313.
- [9] A. McLennan, Consequences of the condorcet jury theorem for beneficial information aggregation by rational agents, *Amer. Polit. Sci. Rev.* 92 (1998) 413–418.
- [10] J.R. Munkres, *Topology: A First Course*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
- [11] I. Obara, Private strategy and efficiency: repeated partnership games revisited, Working Paper, University of Pennsylvania, 2000.
- [12] R. Radner, Repeated partnership games with imperfect monitoring and no discounting, *Rev. Econ. Stud.* 53 (1986) 43–58.
- [13] R. Radner, R. Myerson, E. Maskin, An example of a repeated partnership game with discounting and with uniformly inefficient equilibria, *Rev. Econ. Stud.* 53 (1986) 59–70.
- [14] A. Rubinstein, Why are certain properties of binary relations relatively more common in natural language?, *Econometrica* 64 (1996) 343–355.
- [15] A. Rubinstein, *Economics and Language*, Cambridge University Press, Cambridge, UK, 2000.