Federal Reserve Bank of Minneapolis
Research Department Staff Report 299

February 2002

Learning from Failure*

Andreas Blume

University of Pittsburgh

April Franco

Federal Reserve Bank of Minneapolis
and University of Iowa

ABSTRACT

We study decentralized learning in organizations. Decentralization is captured through a symmetry constraint on agents' strategies. Among such *attainable strategies,* we solve for optimal and equilibrium strategies. We model the organization as a repeated game with imperfectly observable actions. A fixed but unknown subset of action profiles are *successes* and all other action profiles are *failures*. The game is played until either there is a success or the time horizon is reached. For any time horizon, including infinity, we demonstrate existence of optimal attainable strategies and show that they are Nash equilibria. For some time horizons, we can solve explicitly for the optimal attainable strategies and show uniqueness. The solution connects the learning behavior of agents to the fundamentals that characterize the organization: *Agents in the organization respond more slowly to failure as the future becomes more important, the size of the organization increases and the probability of success decreases.*

*We enjoyed talking to Paul Heidhues and Andy McLennan. This version was created January 31, 2002. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

Organizations are groups of agents linked by a common purpose (North [1994]). An organization will be more successful at achieving the common goal, the better its members can coordinate their actions. Institutions like hierarchies and information systems help in solving an organization's coordination problems. But inevitably some of the decision making in an organization will remain decentralized.

Because of decentralization, organizations will behave differently from individuals, certain tasks are more difficult for organizations than for individuals and, more generally, the behavior of an organization may depend on its size. Even in an organization whose production function is known, coordinating actions may be nontrivial if there are multiple optimal combinations. The organization's task is more challenging if the organization has to search for an optimal production plan.

Day and Tinney [1968] study decentralized learning in a firm with two independent decision makers. Their decision makers respond to success and failure by modifying their decision rules until a satisficing criterion is met. They examine the roles of "caution," "daring" and "failure response" and show that "learning - tempered by caution in response to failure" can eventually solve the firm's problem. Milgrom and Roberts [1990a] describe the coordination problem of a team of managers that has to adapt to changing technological opportunities. They show that if the production function is supermodular, adaptive dynamics will find the firm's optimum. Outside of economics, there is considerable interest in concurrent learning in multiagent environments in the computer science literature. For example, Sen and Sekaran [1998] investigate reinforcement learning in multiagent systems.

None of these approaches attempt either to predict the learning rule or to prescribe an optimal rule. Indeed, these questions are not well-posed in these models. Day and Tinney's learning rules are primitives of their model. The computer science literature, e.g., Sen and Sekaran, emphasizes domain independence and robustness. Milgrom and Roberts focus on the outcomes achievable through learning and do not impose restrictions on learning rules that would distinguish "optimal learning" from going straight to the firm's optimum. We impose such constraints and show that the investigation of *optimal learning* in organizations is both meaningful and interesting. In particular, we will be able to say just how "daring" individuals in the organization ought to be and how decisively the they should respond to failure. Moreover, we get to connect these predictions to the fundamental parameters of the model: the discount factor, the size of the organization and the success probability.

We study decentralized learning in organizations. We require agents to learn by trial and error. They can observe only success or failure and not the actions taken by other agents. A failure means that the organizations 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 [1990] (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 but unknown subset of action profiles are *successes* and all other action profiles are *failures.* The game is played until either there is a success or the time horizon, T, is reached.

If agents use expected present discounted values to evaluate outcomes in the repeated game, we find for any time horizon, including infinity, that optimal attainable strategies exist and that they form equilibria. Our central result is a characterization of optimal and equilibrium attainable strategies when each agent has two actions and $T = 3$. We show that there is a unique optimal attainable strategy and that it coincides with the unique equilibrium attainable strategy. This strategy is parameterized by the discount factor, δ, the number of members of the organization, n, and the number of success profiles, k. We find that agents in an organization respond more slowly to failure as the future becomes more important, the size of the organization increases and the probability of success decreases.

If we require sequential rationality, then for $\delta = 0$, $n = 2$ and any T, there exists an optimal equilibrium attainable with an interesting simple structure: Agents alternate between switching with probability one and with probability one-half until either there is a success or the time-horizon is reached. Further, every optimal attainable strategy has this property. All solutions for sufficiently small δ approximate this $\delta = 0$ solution. Finally, in this environment, if we permit some regard for the future by giving agents lexicographic preferences for early success, the optimal attainable strategy remains the same, but no attainable Nash equilibrium exists.

Except when $\delta = 0$, attainable equilibria tend to be complex. The unique attainable equilibrium for $T = 3$ is not in public strategies. More generally, plausible heuristics, like always switching with a fixed probability, do not give rise to equilibria.

2 The Game and Restrictions on Strategies

We consider a repeated n-player, m-action game. In the stage game, the players, $i \in I$, have actions $a_{ij},\ j = 1, \ldots, m$ and identical payoffs from each action combination $a = (a_{1j_1}, \ldots, a_{nj_n})$. All action combinations are either *failures* or *successes*. 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. 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.

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 $t = 0, 1, \ldots, T$, until either a success is played once or the time horizon T is reached.[1] Note that period 0 is the first period; period 1, the second period; and so on. We consider both finite and infinite T. Denote the repeated game with time horizon T by Γ^T. Players observe only their own actions and their own payoffs, not the actions of the other player. We assume that players maximize the expected present discounted value of future payoffs with $0 \leq \delta < 1$; except where indicated, we assume that $\delta > 0$. Thus preferences over strategy profiles have a utility function representation, which is of the von Neumann-Morgenstern type.[2]

Let A_i be player i's set of actions in the stage game and $S = \{0, c\}$ the set of possible

[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.

[2] Toward the end we will also briefly comment on lexicographic preferences for early success.

payoff realizations in the stage game, failure and success. We adopt the convention that after a success, histories record the null event "nothing happens" in each period until the final period T is reached. Since players can observe only success or failure in addition to their own actions, their strategies can depend only on their private histories $h_i^t \in A_i^t \times S^t$. Let $h^t := (h_1^t, ..., h_n^t)$ be an entire history ending in period t, and denote an infinite history by h. Let H stand for the set of all infinite histories.

A (behavior) strategy for player i is a sequence of functions

$$f_i^t : A_i^{t-1} \times S^{t-1} \to \Delta(A_i),$$

where $\Delta(A_i)$ denotes the set of probability distributions over A_i. For any history $h_i^{t-1} \in A_i^{t-1} \times S^{t-1}$ and $a \in A_i$, let $f_i^t(h_i^{t-1})(a)$ denote the probability that player i's strategy assigns to his action a after history h_i^{t-1}.

Doing well in this game means that players quickly find a success. They will want to avoid action combinations that have led to failures in the past. They will want to learn from their failures. With only a single player, someone directing the organization, or some other mechanism coordinating the organization, this is accomplished by trying a new action combination in each period until there is a success. With decentralized decision making by two agents this search process is made difficult by the fact that agents won't know who took the wrong action: agent 1, agent 2, or both. With more than two agents the difficulty increases because any subset of the set of agents could have taken a wrong action.

We use symmetry to capture the non-negligible strategic uncertainty that players are facing. In particular, since we do not permit prior communication or any other prior interaction, symmetries of the game can be removed only through interactions *in* the game. As a consequence, 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 defined only up to a permutation of A_i. Formally, these two conditions amount to

(1) $\exists$ a bijection $\hat{\rho} : A_i \to A_{i'}$ such that $f_i^t(h_i^{t-1})(a) = f_{i'}^t(\hat{\rho}(h_i^{t-1}))(\hat{\rho}(a))$, and

(2) $f_i^t(h_i^{t-1})(a) = f_i^t(\tilde{\rho}(h_i^{t-1}))(\tilde{\rho}(a))$, $\forall$ bijections $\tilde{\rho} : A_i \to A_i$.

We call strategy profiles that satisfy our restrictions *attainable.* Strategy profiles that maximize players' payoffs among the set of attainable strategies will be called *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. Our goal is to characterize optimal attainable strategy profiles and equilibrium attainable strategy profiles in repeated success-or-failure games. Individual strategies will be called attainable if they satisfy the second condition. 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 success-or-failure game by $\mathcal{F}_i^T$ and the set of attainable strategy profiles by $\mathcal{F}^T$.

Since payoffs in the stage game are bounded and future payoffs are discounted with discount factor $\delta < 1$, the repeated success-or-failure game is *continuous at infinity* (see Fudenberg and Levine [1983]), 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

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

This continuity property will be used in the next section to help establish existence of optimal attainable strategies and of equilibrium attainable strategies.

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 $\pi_t(f^0, \ldots, f^t)$, that is, as a function of behavior only up to and including period t. In the discounting case, we can then write each player's payoff from profile f as

$$\Pi(f) = \sum_{t=0}^{T} \delta^t \pi_t(f^0, \ldots, f^t).$$

For any two strategies f and g denote by $g(f,\tau)$ the strategy obtained from g by following f until period τ and g thereafter. Then, as a consequence of continuity at infinity we have the following property: For any $\epsilon > 0$, there exists τ_0 such that

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

3 Optimal Attainable Strategies: Existence and Relation to Equilibrium

In this section we establish the existence of optimal attainable strategies in repeated success-or-failure games for any time horizon and show that optimal attainable strategies are Nash equilibrium strategies. This conveniently establishes the existence of equilibrium attainable strategies.[3]

We begin by proving existence of optimal attainable strategies. The idea is simple: We have to show that a constrained optimization problem has a solution. This will be the case if the objective function and constraint set are well behaved.

Proposition 1 *An optimal attainable strategy exists in any repeated success-or-failure game with discounting.*

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 a product of the spaces of 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.[4]

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.

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

[3]While the equilibrium property of optimal attainable strategies is convenient and intuitive, it deserves emphasis that this link may fail, despite the common interest among players, if players use criteria other than expected present discounted values to evaluate outcomes in the repeated game. An interesting case where this occurs is that of lexicographic preferences for early success.

[4]For this and later references to topology see, e.g., Munkres [1975].

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 $\mathcal{F}^\infty$ 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 $\epsilon > 0$ such that $\Pi(\hat{f}) - \Pi(\tilde{f}) > \epsilon$. 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

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

$\Pi(\tilde{f}_n^*)$ 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 $|\Pi(\tilde{f}_T) - \Pi(\tilde{f}_T^*)| < \frac{\epsilon}{4}$.

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

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. □

When players have identical payoffs, as in the repeated success-or-failure game, it is intuitive that optimal attainable strategy combinations are Nash equilibria. The following result confirms this intuition.

Proposition 2 *Any optimal attainable strategy in a repeated success-or-failure game with discounting is a Nash equilibrium strategy.*[5]

[5]CH prove an analogous result for repeated coordination games without payoff uncertainty.

The proof makes use of the fact that preferences over strategy profiles have utility function representation and, more specifically, that this representation has the von Neumann-Morgenstern form. We will show later that without this assumption there may be no attainable Nash equilibrium in the repeated success-or-failure game, even if we permit the use of mixed strategies.

Proof. 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 strategy $\hat{f}_1$ for player one such that

$$\Pi(\hat{f}_1, f^*_{-1}) > \Pi(f^*).$$

Note that because the game is symmetric, it is without loss of generality to consider only deviations of player one. Since actions are payoff equivalent as long as they have not been taken, it is also without loss of generality to let $\hat{f}_1$ be an attainable strategy. Consider a strategy $\tilde{f}$ in which each player plays f^* with probability $1-\epsilon$ and $\hat{f}$ with probability ϵ. Denote the common payoff from m players using strategy $\hat{f}$ and the remaining $n-m$ players using strategy f^* by $\Pi^{m,n}(\hat{f}; f^*)$; this uses the fact that preferences over strategy profiles have a utility-function representation. Then the common expected payoff from $\tilde{f}$ is

$$\Pi(\tilde{f}) = \sum_{m=0}^{n} \tbinom{n}{m} \epsilon^m (1-\epsilon)^{n-m} \Pi^{m,n}(\hat{f}; f^*)$$

$$= (1-\epsilon)^n \Pi(f^*) + n\epsilon(1-\epsilon)^{n-1}\Pi(\hat{f}_1, f^*_{-1}) + \sum_{m=2}^{n} \tbinom{n}{m} \epsilon^m (1-\epsilon)^{n-m} \Pi^{m,n}(\hat{f}; f^*).$$

Here, we use the fact that the utility function representing preferences has the von Neumann-Morgenstern form. 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. □

Combining propositions 1 and 2, we immediately obtain the following existence result for attainable equilibria.

Corollary 1 *An optimal attainable Nash equilibrium strategy exists in any repeated success-or-failure game.*

4 The Solution for the Three-Period Case

In this section, we study the three-period game with two actions per player, n players and k success profiles under the assumption that the n players discount the future with an identical constant discount factor δ. We assume that the number of success locations is less than the total number of n-tuples available to the players minus 1, i.e., $k < 2^n - 1$, to ensure that an individual player cannot guarantee a success in the three-period game.

For tractability reasons, here and in the following sections, we confine ourselves to the case where players each have two actions. In that case, our two attainability conditions amount to players using identical switching probabilities, given identical switching histories. Let a "1" denote a switch from one action to another and a "0" no such switch. Define $W := \{0, 1\}$. Then a player's "relevant history" at time $t = 1, 2, \dots$ is an element of $W^{t-1} \times S^{t-1}$, the history of his own switches and of the (publicly observable) failures and successes. We can analyze the game in terms of players' "switching strategies":

$$\sigma_i : W^{t-1} \times S^{t-1} \to \Delta(W).$$

Our condition (2) for attainability implies that players use switching strategies. Condition (1) implies that these switching strategies have to be identical. We will therefore study pairs of identical switching strategies and see which ones are optimal and which ones are in Nash equilibrium.

Since it suffices to study switching strategies in the two-action case and since the initial choice probabilities are determined by the attainability constraint, 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 one, which is the first period in which players can switch, q_0 is the probability of switching in period two conditional on no switch in the previous period, and q_1 is the probability of switching in period two conditional on a switch in the previous period.

The following result establishes the existence of a *unique* optimal attainable strategy. The proof is constructive and determines the optimal values of p, q_0 and q_1 as functions of the discount factor, δ, the number of players, n, and the number of success profiles, k.

Proposition 3 *In the three-period repeated failure-or-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* $(p, q_0, q_1) = \left(p^*(\delta, n, k), 1, \frac{1}{2}\right)$*, where*

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

Proof. If the n players use the switching strategy (p, q_0, q_1), each player's expected payoff is given by

$$\pi(p, q_0, q_1) = c\frac{k}{2^n} + c\left(\frac{2^n - k}{2^n}\right)\delta \left(\begin{array}{c} \frac{k}{2^n-1}\left(1-(1-p)^n\right) \\ +\delta\left(\begin{array}{c} (1-p)^n \frac{k}{2^n-1}\left(1-(1-q_0)^n\right) \\ +\Sigma_{j=1}^n p^j (1-p)^{n-j}\binom{n}{j} \\ \left(\frac{2^n-1-k}{2^n-1}\right)\left(\frac{k}{2^n-2}\right)\left(\begin{array}{c} 1 \\ -q_1^j (1-q_0)^{n-j} \\ -(1-q_1)^j (1-q_0)^{n-j} \end{array}\right) \end{array}\right) \end{array}\right).$$

Maximizing $\pi(p, q_0, q_1)$ is equivalent to maximizing

$$\tilde{\pi}(p, q_0, q_1) = \left(\frac{2^n - k}{2^n}\right) \left(\begin{array}{c} \left(1-(1-p)^n\right) \\ +\delta\left(\begin{array}{c} (1-p)^n\left(1-(1-q_0)^n\right) \\ +\Sigma_{j=1}^n p^j (1-p)^{n-j}\binom{n}{j} \\ \left(\frac{2^n-1-k}{2^n-2}\right)\left(\begin{array}{c} 1 - q_1^j (1-q_0)^{n-j} \\ -(1-q_1)^j (1-q_0)^{n-j} \end{array}\right) \end{array}\right) \end{array}\right).$$

For $p = 0$, the conditionally optimal value of q_0 is 1, giving $\tilde{\pi}(0, 1, q_1) = \frac{2^n-k}{2^n}\delta$. In contrast, for $p = 1$, $\tilde{\pi}(1, q_0, q_1) \geq \frac{2^n-k}{2^n}$. This implies that $p = 0$ cannot be part of an optimal strategy.

Since $p > 0$, the partial derivative of $\tilde{\pi}(1, q_0, q_1)$ with respect to q_1 is

$$\delta\Sigma_{j=1}^n p^j (1-p)^{n-j}\binom{n}{j}\left(\frac{2^n - 1 - k}{2^n - 2}\right)\left(\begin{array}{c} -jq_1^{j-1}(1-q_0)^{n-j} \\ +j(1-q_1)^{j-1}(1-q_0)^{n-j} \end{array}\right).$$

Hence, the corresponding first order condition has a unique solution, $q_1 = \frac{1}{2}$. One also easily checks that the second derivative is negative.

Next, we consider the optimal solution for q_0. If $p \in (0,1)$, then the partial derivative of $\tilde{\pi}(1, q_0, q_1)$ with respect to q_0 is

$$\delta\left(\begin{array}{c} (1-p)^n\left(n(1-q_0)^{n-1}\right) \\ +\Sigma_{j=1}^n p^j (1-p)^{n-j}\binom{n}{j}\left(\frac{2^n-1-k}{2^n-2}\right)\left(\begin{array}{c} (n-j)\, q_1^j (1-q_0)^{n-j-1} \\ +(n-j)(1-q_1)^j (1-q_0)^{n-j-1} \end{array}\right) \end{array}\right).$$

In the case where $p \in (0,1)$, this derivative is positive for all values of q_0 that are less than one. Hence, in that case the optimal value of q_0 is 1. If $p = 1$, then q_0 is unrestricted. Hence, the optimal value satisfies $q_0^* = 1$ if we can show that $p^* \in (0,1)$. This is the question to which we turn next.

Finally, we consider the solution for p. Using the solutions to q_0 and q_1, the partial derivative of $\tilde{\pi}(1, q_0, q_1)$ with respect to p is

$$\left(\frac{2^n-k}{2^n}\right)\left[\begin{array}{c} n(1-p)^{n-1} - n\delta(1-p)^{n-1} \\ +\delta\left(\frac{2^n-1-k}{2^n-2}\right)\left\{\begin{array}{c} \Sigma_{j=1}^{n-1}\left(\begin{array}{c} jp^{j-1}(1-p)^{n-j}\binom{n}{j} \\ -(n-j)\, p^j (1-p)^{n-j-1}\binom{n}{j} \end{array}\right) \\ +np^{n-1}\left(1-q_1^n-(1-q_1)^n\right) \end{array}\right\} \end{array}\right].$$

Because

$$(n-j)\binom{n}{j} = (j+1)\binom{n}{j+1}$$

this is equivalent to

$$\left(\frac{2^n-k}{2^n}\right)\left[\begin{array}{c} n(1-p)^{n-1} - n\delta\left(\left(1-\left(\frac{2^n-1-k}{2^n-2}\right)\right)(1-p)^{n-1}\right) \\ -\delta\left(\frac{2^n-1-k}{2^n-2}\right)np^{n-1} + \delta\left(\frac{2^n-1-k}{2^n-2}\right)np^{n-1}\left(1-q_1^n-(1-q_1)^n\right) \end{array}\right].$$

One easily checks that the second derivative is negative for all feasible values of p. Hence, a solution to the first-order condition is a global optimum. The first-order condition implies

$$\begin{aligned} &((1-p)^{n-1} - \delta(1-p)^{n-1}) \\ &\quad +\delta(1-p)^{n-1}\frac{2^n-1-k}{2^n-2} - \delta p^{n-1}\frac{2^n-1-k}{2^n-2}(q_1^n + (1-q_1)^n) = 0. \end{aligned}$$

After dividing by p^n and solving the resulting equation for $\left(\frac{1-p}{p}\right)^{n-1}$, we get

$$\left(\frac{1-p}{p}\right)^{n-1} = \frac{\delta(2^n-1-k)(q_1^n + (1-q_1)^n)}{(1-\delta)(2^n-2) + \delta(2^n-1-k)}.$$

After taking the $(n-1)$th root, using $q_1 = \frac{1}{2}$ and solving for p, we are left with

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

as claimed. $\square$

It is interesting to compare this solution for the three-period game with the optimal attainable strategy for the two-period game. In the two period-game, we only have to determine the switching probability, p, in the second period. Since we want to minimize the probability $(1-p)^2$ that the first-period action profile is revisited, $p = 1$ is uniquely optimal. In contrast, in the three-period game it is optimal to sacrifice some second-period success probability in order to improve the probability of a better starting position in the third period.

Since every optimal attainable strategy is an equilibrium strategy, the strategy identified in the last result is an attainable equilibrium strategy. Since it is unique, it is also the unique optimal attainable equilibrium strategy. The following result shows that uniqueness is preserved even if we drop the optimality requirement. There is a unique attainable equilibrium strategy.

Proposition 4 *In the three-period repeated failure-or-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 is given by $(p, q_0, q_1) = \left(p^*(\delta, n, k), 1, \frac{1}{2}\right)$ and thus coincides with the optimal attainable strategy.*

Proof. First consider the value of p. Suppose that $p = 1$. Then player i can guarantee that in the final period a new action combination is used by defecting to the strategy $(0, q_0 = 1, q_1)$. Hence $p = 1$ is never a part of an attainable equilibrium strategy. Suppose instead that $p = 0$. Player i gains by deviating to $(p = 1, q_0, q_1)$, which raises the probability of using a new action combination in period one from zero to one, without affecting the probability of a new action combination in the final period. Hence, $p = 0$ is never part of the attainable equilibrium strategy.

So, in any attainable equilibrium strategy $p \in (0, 1)$. Since there is positive probability that at least one player i does not switch in the second period, at least one other player j does switch. This implies that it is uniquely optimal for player i to set $q_0^i = 1$.

Since there is positive probability that player i switches in the second period, q_1^i must be optimal in the third period. Recall that by symmetry, $p^{j\neq i} = p^i$ for all j. Against (p, q_0, q_1), the payoff in the third period from switching after having switched in the previous period is

$$c\left(\frac{k}{2^n-2}\right)\Sigma_{j=1}^{n-1} p^{j-1}(1-p)^{n-j}\tbinom{n-1}{j-1}\left[1-q_1^{j-1}(1-q_0)^{n-j}\right]$$

while the expected payoff from not switching after having switched in the previous period is

$$c\left(\frac{k}{2^n-2}\right)\Sigma_{j=1}^{n-1} p^{j-1}(1-p)^{n-j}\tbinom{n-1}{j-1}\left[1-(1-q_1)^{n-j}(1-q_0)^{n-j}\right].$$

Consequently, if q_1 is greater than $\frac{1}{2}$, it is uniquely optimal not to switch in period two following a switch in period one, i.e., to deviate to $q_1 = 0$. Conversely, if q_1 is less than $\frac{1}{2}$, it is uniquely optimal to switch in period two following a switch in period one, i.e. to deviate to $q_1 = 1$. Thus, the only possibility for q_1 to be optimal is that $q_1 = \frac{1}{2}$. Indeed, $q_1 = \frac{1}{2}$ equates the payoffs from switching and not switching following a switch.

Since $p \in (0,1)$ in any attainable equilibrium strategy, player i's payoff from switching in the second period must be identical to the payoff from not switching. Player i's payoff from not switching against the putative equilibrium strategy $\left(p, 1, \frac{1}{2}\right)$ in the second period, $\Pi^i\left(s^i = \left(0, 1, \frac{1}{2}\right), s^{j\neq i} = \left(p, 1, \frac{1}{2}\right)\right)$, is

$$c\frac{k}{2^n} + c\left(\frac{2^n-k}{2^n}\right)\delta\left(\begin{array}{c} \frac{k}{2^n-1}\left(\begin{array}{c} 1-(1-p)^{n-1} \\ +\delta(1-p)^{n-1}\end{array}\right) \\ +\delta\Sigma_{j=1}^{n-1}\left\{\begin{array}{c} p^j(1-p)^{n-j-1}\binom{n-1}{j}\left(\frac{2^n-1-k}{2^n-1}\right) \\ \left(\frac{k}{2^n-2}\right)\left(\begin{array}{c} 1-q_1^j(1-q_0)^{n-j} \\ -(1-q_1)^j(1-q_0)^{n-j}\end{array}\right)\end{array}\right\}\end{array}\right).$$

Player i's expected payoff from switching against the putative equilibrium strategy $\left(p, 1, \frac{1}{2}\right)$ used by players $j \neq i$ in the second period, $\Pi^i\left(s^i = \left(1, 1, \frac{1}{2}\right), s^{j\neq i} = \left(p, 1, \frac{1}{2}\right)\right)$, is

$$c\frac{k}{2^n} + c\left(\frac{2^n-k}{2^n}\right)\delta\left(\begin{array}{c} \frac{k}{2^n-1} \\ +\delta\Sigma_{j=1}^{n}\left\{\begin{array}{c} p^{j-1}(1-p)^{n-j}\left(\frac{2^n-1-k}{2^n-1}\right) \\ \left(\frac{k}{2^n-2}\right)\left(\begin{array}{c} 1-q_1^j(1-q_0)^{n-j} \\ -(1-q_1)^j(1-q_0)^{n-j}\end{array}\right)\end{array}\right\}\end{array}\right).$$

Equating the payoffs from switching in the second period and from not switching in the second period and using the fact that $q_0 = 1$, we get

$$(-(1-p)^{n-1}+\delta(1-p)^{n-1})\frac{2^n-2}{2^n-1}+\delta\sum_{j=1}^{n-1}p^j(1-p)^{n-j-1}\left(\begin{array}{c}n-1\\j\end{array}\right)\frac{2^n-1-k}{2^n-1}$$

$$=\delta\sum_{j=1}^{n-1}p^{j-1}(1-p)^{n-j}\left(\begin{array}{c}n-1\\j-1\end{array}\right)\frac{2^n-1-k}{2^n-1}+\delta p^{n-1}\frac{2^n-1-k}{2^n-1}(1-q_1^n-(1-q_1)^n).$$

After eliminating matching terms on both sides of the equation, we are left with

$$(-(1-p)^{n-1}+\delta(1-p)^{n-1})\frac{2^n-2}{2^n-1}$$

$$=\delta(1-p)^{n-1}\frac{2^n-1-k}{2^n-1}-\delta p^{n-1}\frac{2^n-1-k}{2^n-1}(q_1^n+(1-q_1)^n).$$

Now, one can easily check that this equation is equivalent to the first-order condition that determined $p^*(\delta,n,k)$ as part of the unique optimal attainable strategy. □

Note that in the unique attainable Nash equilibrium, agents condition their behavior on their own past actions, which are not publicly observable. Therefore no public attainable Nash equilibrium exists in our game. There has been some recent work on the effects of restricting the analysis of games with imperfect monitoring to public equilibria. Fudenberg, Levine and Maskin [1994] 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, Myerson and Maskin [1986] 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 [1986] shows that efficiency can be attained in repeated partnership games without discounting. Recently Obara [2000] 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.

Simple inspection of $p^*(\delta,n,k)$ yields a number of interesting comparative statics predictions.

Proposition 5 *In the three-period repeated failure-or-success game with two actions per player, common discount factor δ, n players, and k success profiles, the unique optimal and equilibrium value of the second-period switching probability p is a monotonic function of δ, n and k. The function $p^*(\delta, n, k)$ (1) is strictly decreasing in δ, (2) is strictly decreasing in n, (3) is strictly decreasing in n even if $\frac{k}{2^n}$ is kept constant, (4) is strictly increasing in k, (5) converges to $\frac{2}{3}$ as $\delta \to 1$, (6) converges to 1 as $\delta \to 0$, and (7) converges to $\frac{2}{3}$ as $n \to \infty$.*

For the three-period case, the proposition summarizes how agents' responses to failure in an organization vary with the fundamentals that characterize the organization. Since the third-period switching probabilities do not vary with the parameters, the proposition focuses on the switching probability in the second period, following a failure in the initial period. In the three-period model all three parameters, i.e., the discount factor, the number of agents in the organization, and the initial success probability, affect this switching probability; this contrasts with the two-period case where agents always prefer to switch with probability one following a failure.[6] We infer the following from the proposition: (1) Agents in the organization respond more slowly to failure as the future becomes more important. Rather than always immediately seizing a new success opportunity, agents take into account how their behavior affects the point of departure in the next period if they do not succeed in the present period. (2) Agents respond more slowly to failure as the size of the organization increases. The larger the organization, the greater is the incentive to rely on others for change. Note that the benefit from relying on others does not derive from free-riding but from the increased probability of generating asymmetries. (3) The effect of increasing organization size on agents' responses to failure remains the same even if the *a priori* success probability is kept constant. This can be interpreted as meaning that the size effect remains even if large organizations have the same technology as small ones. Since the effect of size on reactivity is strictly negative in this case, it would remain so even if larger organizations had slightly better technologies. (4) Agents in the organization respond faster to failure as the success probability is increased. (5) Even very patient agents change their behavior with positive

[6] It is difficult to obtain analytical and even numerical solutions for four and more periods. We have explored the four-period case for $n = 2$ and $k = 1$. In that case, the comparative statics for δ remain the same as in the three-period case. We expect that at least for moderate discount factors, adding more periods does not affect our comparative statics results too much. For small discount factors, $n = 2$ and $k = 1$, the infinite horizon case is analyzed in the following section.

probability. (6) Very impatient agents seek immediate success. (7) Even in a very large organization agents respond to failure with a probability that is bounded away from zero.

5 The Infinite Horizon Case

For large T, the analysis of the general version of the repeated success-and-failure game becomes increasingly intractable, even if we limit ourselves to two players, two actions and a single success pair, as we will from here on. It turns out, however, that for small discount factors we can approximately characterize the solution for any horizon. The case of small discount factors is interesting because, unlike in the case of ordinary repeated games, there are significant intertemporal links no matter how small the discount factor. Indeed, the analysis of the limit lays bare the role of information in determining play.

For expositional purposes, we formulate our results for the infinite horizon. Statements and proofs apply to any horizon with hardly any changes. Our approach will be to identify the solution for $\delta = 0$ and then to argue that any solution for small but positive δ approximates the solution for $\delta = 0$. Some care is required in order to discuss in a meaningful way optimal attainable strategies when $\delta = 0$. In particular, we want to get restrictions on behavior not just in the first period. For that reason we employ a mild *sequential optimality* condition: We require an optimal attainable strategy to remain optimal attainable after every positive probability history. The same holds for equilibrium attainable strategies. For $\delta > 0$, this condition is automatically satisfied. Thus, it is natural to maintain it in studying optimal solutions in the limit.

To state the next result, we need to introduce switching probabilities conditional on sets of histories. By definition, a switching strategy σ specifies a switching probability $\sigma(h^t)$ conditional on each individual history h^t ending in period $t-1$. In addition, for any set of histories $H^{\alpha t}$ ending in period $t-1$ that has positive probability under σ we can define an aggregate switching probability, the probability of switching conditional on the history being in the set $H^{\alpha t}$; α is an index that is either "odd" or "even" and will later be used to differentiate between the set of length $t-1$ histories with an even number and an odd number of switches. Let $\rho(h^t, \sigma)$ denote the probability of history h^t that is induced by switching strategy σ, and define the probability of history h^t conditional on being in the

σ-positive-probability set $H^{\alpha t}$ as

$$\rho(h^t, \sigma | H^{\alpha t}) := \frac{\rho(h^t, \sigma)}{\sum_{h'^t \in H^{\alpha t}} \rho(h'^t, \sigma)}.$$

The switching probability conditional on set $H^{\alpha t}$ can then be defined as

$$\sigma(H^{\alpha t}) := \sum_{h^t \in H^{\alpha t}} \sigma(h^t) \rho(h^t, \sigma | H^{\alpha t}).$$

Note that for any σ-positive-probability set $H^{\alpha t}$ the switching probability conditional on $H^{\alpha t}$ is independent of the specification of switching probabilities after individual histories that are in $H^{\alpha t}$ but have probability zero under σ. Define the set H^{et} of histories with an even number of switches and the set H^{ot} of histories with an odd number of switches, and call the switching probabilities conditional on these sets, when they are well-defined, the switching probability conditional on even and odd histories, respectively.

Proposition 6 *In the infinite horizon repeated failure-or-success game with two actions per player, common discount factor $\delta = 0$, 2 players, and 1 success profile, (1) there exists a public optimal attainable strategy σ^* that prescribes switching with probability one in odd periods and with probability one-half in even periods; and, (2) in every optimal attainable strategy σ the switching probability is one in odd periods, and in even periods both the switching probability conditional on odd histories and the switching probability conditional on even histories equal one-half.*

Proof. We show that any attainable strategy σ whose switching probabilities conditional on odd and even histories equal one-half in even periods and whose switching probability in odd periods equals one is an optimal attainable strategy. We proceed by induction. We first show that switching with probability one is uniquely optimal in period one and that conditional on having switched with probability one in period one, it is uniquely optimal to switch with probability one-half in period two. We then show for any t that if σ is as specified until period $t-1$ then it is uniquely optimal to satisfy the specification in period t.

Recall that the initial, 0 period, randomization is constrained to assigning probability one-half to either action. If the 0-period-action pair was not successful, switching with probability one is the only way to ensure that the same pair will not be chosen in period

1. Since all other action pairs are equally likely to be the success pair, minimizing the probability of repeating the first-round choices is *uniquely* optimal in period one. If there is a unique optimal switching probability in a given period, $\delta = 0$ requires that this switching probability be part of any optimal strategy.

Without success prior to period 2, the goal in period 2 is to minimize the probability of repeating the two first-round choice pairs. A new action pair will be chosen exactly when one player switches and the other does not. Denoting the switching probability in period 2 by σ_2, the probability of success in that period is $\frac{1}{2}\sigma_2(1-\sigma_2)$, which is again *uniquely* maximized with a switching probability of one-half.

A similar argument applies after any success-free history conforming with σ in which the last switch was a probability one switch. The argument is slightly different because we have to allow players to condition their switching probabilities on their individual histories.

Let period t be even. Since period t is even, by assumption both players switched with probability one in the preceding period. Since we are in period four or higher, there have been at least two simultaneous probability-one switches. Hence it is impossible to have reached the current period without success while one player's history is even and the other player's history is odd. Thus, either both players' histories are even, or they are both odd. Therefore, given that σ is as specified in the proposition until period $t-1$, with probability one-half both players' histories are in H^{et}, and otherwise they are both in H^{ot}. In either case, in order to have a chance for a success one of the players has to switch while the other stays put. There are two ways in which this can happen, and if it does there is a one-half chance of a success. Therefore, the success probability conditional on having reached period t is

$$\Pi(t) = c\frac{1}{2}2\sigma(H^{et})(1-\sigma(H^{et}))\frac{1}{2} + c\frac{1}{2}2\sigma(H^{ot})(1-\sigma(H^{ot}))\frac{1}{2} =$$

$$c\frac{1}{2}\sigma(H^{et})(1-\sigma(H^{et})) + c\frac{1}{2}\sigma(H^{ot})(1-\sigma(H^{ot})).$$

This function attains its maximum whenever

$$\sigma(H^{et}) = \sigma(H^{ot}) = \frac{1}{2}.$$

It remains to show that if t is odd and given that σ has been followed until period $t-1$, it is uniquely optimal to switch with probability one in period t. By assumption, both players switched with probability one-half in period $t-1$ and with probability one in period $t-2$. Conditional on no success in period $t-1$, a precondition for arriving in period t, the probability of being at one of the two action pairs visited in periods $t-3$ and $t-2$ is two-thirds and the probability of being at one of the remaining action pairs is one-third. In the two-thirds case, for a success to be possible one of the players must switch while the other stays put. With probability one-half this happens when both players' histories are even (odd), in which case there are two ways in which it can happen, both of which have success probability one-half. In the one-third case, there is certain success if and only if both players switch. Hence, if t is odd and σ has been followed until period $t-1$, then the success probability in period t is

$$\Pi(t) = c\frac{2}{3}(\frac{1}{2}\sigma(H^{et})(1-\sigma(H^{et})) + c\frac{1}{2}\sigma(H^{ot})(1-\sigma(H^{ot}))) + c\frac{1}{3}\sigma(H^{et})\sigma(H^{ot}).$$

The derivative of the success probability with respect to $\sigma(H^{et})$ is

$$\frac{\partial\Pi(t)}{\partial\sigma(H^{et})} = c\frac{1}{3} - c\frac{2}{3}\sigma(H^{et}) + c\frac{1}{3}\sigma(H^{ot}),$$

and similarly

$$\frac{\partial\Pi(t)}{\partial\sigma(H^{ot})} = c\frac{1}{3} - c\frac{2}{3}\sigma(H^{ot}) + c\frac{1}{3}\sigma(H^{et}).$$

Either both switching probabilities equal one, and we are done, or the sum of both derivatives is positive. Suppose the latter holds. Since the sum of the two expressions is positive, at least one of them must be positive. Suppose $\frac{\partial\Pi(t)}{\partial\sigma(H^{et})} > 0$ at the optimum, then $\sigma(H^{et}) = 1$. But then $\frac{\partial\Pi(t)}{\partial\sigma(H^{ot})} = \frac{2}{3} - \frac{2}{3}\sigma(H^{ot})$, which is strictly greater than zero unless $\sigma(H^{ot}) = 1$ as well. Thus, in either case $\sigma(H^{ot}) = 1$. Since the argument is symmetric, it has to be the case that $\sigma(h) = 1\ \forall h \in H^{et}$ and $\sigma(h) = 1\ \forall h \in H^{ot}$. □

The optimal attainable strategy of switching in odd periods and randomizing with probability one-half in even periods conditions only on time and is therefore a public strategy. This contrasts with the case of $\delta > 0$ and three-periods, where the unique optimal attainable strategy made behavior-contingent on players' own past switches.

Proposition 7 *If we let $\sigma(H^{et};\delta)$ denote the switching probability after even histories under any strategy σ that is optimal attainable given the discount factor δ (and similarly for odd histories), then*

$$\sigma(H^{\alpha t};\delta_n) \to \sigma(H^{\alpha t};0)$$

for $\alpha = e, o$, $t = 1, 2, \ldots$ and $\delta_n \to 0$.

Proof. A moment's reflection shows that the payoff function Π has the following general form:

$$\Pi(\sigma,\delta) = c\frac{1}{4} + c\sum_{t=1}^{\infty} \delta^t \pi_t(\sigma_1,\ldots,\sigma_t),$$

where σ_t is the vector of history-dependent switching probabilities in period t and $\pi_t(\ldots)$ is the probability of success in period t as a function of all probability choices that affect period t, i.e., all those prior to and in period t. Each $\pi_t(\ldots)$ is a polynomial and therefore continuous.

We proceed by induction on t. Recall from the previous result that $\sigma(H^{\alpha 1};0)$ is uniquely determined in any solution to the problem of maximizing $\pi_1(\sigma_1)$. In other words, for any alternative strategy $\tilde{\sigma}$ such that $\tilde{\sigma}(H^{\alpha 1};0) \neq \sigma(H^{\alpha 1};0)$, we have

$$\pi_1(\sigma_1) - \pi_1(\tilde{\sigma}_1) > \epsilon$$

for some $\epsilon > 0$. Therefore

$$\Pi(\sigma,\delta) - \Pi(\tilde{\sigma},\delta) > \delta\epsilon - \delta^2,$$

where we use the fact that the maximum payoff in the game is one, corresponding to immediate success, and that payoffs are bounded below by zero. Hence, as soon as $\delta < \epsilon$, we have

$$\Pi(\sigma,\delta) - \Pi(\tilde{\sigma},\delta) > 0,$$

which establishes our claim for $t = 1$.

Suppose the claim holds for any $\tau < t$. We will show that it holds for period t. We know from the previous result that if $\sigma(H^{\alpha\tau};0) = 1$ for $\alpha = e, o$ and all odd $\tau < t$, and $\sigma(H^{\alpha\tau};0) = \frac{1}{2}$ for $\alpha = e, o$ and all even $\tau < t$, then any attainable σ_t that maximizes $\pi_t(\sigma_1,\ldots,\sigma_{t-1},\sigma_t)$ must satisfy the same two conditions with t replacing τ.

Consider a sequence $\{\sigma(\delta_n)\}$ of optimal attainable strategies corresponding to a sequence of discount factors $\delta_n \to 0$. Suppose, in order to obtain a contradiction, that

$$\sigma(H^{\alpha t}; \delta_n) \not\to \sigma(H^{\alpha t}; 0).$$

Then there exists an η and a subsequence $\{\delta_{n_k}\}$ of $\{\delta_n\}$ such that

$$|\sigma(H^{\alpha t}; \delta_{n_k}) - \sigma(H^{\alpha t}; 0)| > \eta \quad \forall k = 1, 2, \ldots.$$

Then $\sigma(\delta_{n_k})$ has a convergent subsequence, which after reindexing we also denote by $\{\sigma(\delta_{n_k})\}$. Denoting the limit of this subsequence by $\tilde{\sigma}$, we infer that

$$|\tilde{\sigma}(H^{\alpha t}) - \sigma(H^{\alpha t}; 0)| > \eta.$$

From the previous result, there exists an $\epsilon > 0$ such that

$$\pi(\sigma_1(0), \ldots, \sigma_{t-1}(0), \sigma_t(0)) - \pi(\sigma_1(0), \ldots, \sigma_{t-1}(0), \tilde{\sigma}_t) > \epsilon.$$

Thus, using the fact that

$$\tilde{\sigma}(H^{\alpha\tau}) = \sigma(H^{\alpha\tau}; 0) \quad \forall\ \tau < t,$$

we get

$$\Pi(\sigma(0), \delta) - \Pi(\tilde{\sigma}, \delta) > \epsilon\delta^t - \delta^{t+1}.$$

Therefore, for any $\delta < \epsilon$, we get

$$\Pi(\sigma(0), \delta) - \Pi(\tilde{\sigma}, \delta) > 0$$

such that convergence of $\sigma(\delta_{n_k})$ to $\tilde{\sigma}$ and the continuity of Π imply that for all sufficiently large k

$$\Pi(\sigma(0), \delta_{n_k}) - \Pi(\sigma(\delta_{n_k}), \delta_{n_k}) > 0$$

which contradicts the presumed optimality of $\sigma(\delta_{n_k})$. □

Since we proved earlier that every optimal attainable strategy is an equilibrium strategy, we immediately infer that the simple public strategy that is optimal attainable for $\delta = 0$ is also an equilibrium strategy.

Corollary 2 *With an infinite time horizon and $\delta = 0$, there exists an equilibrium attainable strategy in which both players switch with probability one in even periods and with probability one-half in odd periods.*

In the discounting case, there is a simple intuitive link between optimality and equilibrium for attainable strategies. Lest the reader think this is obvious we note that this link disappears if we replace discounting with $\delta = 0$ by a lexicographic preference for early success.

For any two sequences of expected stage-game payoffs of player i, π^i and $\hat{\pi}^i$, define

$$t^* := \max\{t \geq 0 | \pi^i_\tau = \hat{\pi}^i_\tau \ \forall \tau < t\}.$$

We say that player i lexicographically prefers the sequence π^i of expected stage-game payoffs to the sequence $\hat{\pi}^i$, $\pi^i \succ_i \hat{\pi}^i$, if and only if $\pi^i_{t^*} > \hat{\pi}^i_{t^*}$.

Proposition 8 *With lexicographic preferences for early success, the infinitely repeated game does not have an attainable Nash equilibrium.*

Proof. To derive a contradiction, suppose there is such an equilibrium. The attainability restrictions on strategies imply that in the initial (0 period) players uniformly randomize over their actions. If players' switching probability in period one is less than one, each player's unique best reply is to switch with probability one. Thus, the only candidate for an attainable equilibrium requires players to switch with probability one in period one. Then, the switching probability in period two cannot equal zero or one because otherwise each player could gain by deviating to switching with probability one or zero.

Thus in any candidate σ^* for an attainable Nash equilibrium, the period-one switching probability is one and the period-two switching probability σ^*_2 (conditional on switching in period 1) satisfies $0 < \sigma^*_2 < 1$. As a consequence, conditional on no success in the first two periods, the expected number of action combinations examined in the first three periods is less than three. However, a player can deviate to the following strategy: switch with probability zero in period one and with probability one in period two. The success probabilities in periods zero and one are identical to the equilibrium candidate's success probability, and conditional on no success, the number of action combinations examined in the first three periods equal three. $\square$

6 Related Literature

Interest in organizational learning dates back to at least Cyert and March [1963]. They view an organization as an "information gathering and decision-rendering system" and share with Simon [1947] the emphasis on decision processes in organizations. According to them, some problems are routine and handled by "standard operating procedures." If problems change, routines are adjusted through learning. Organizations search for better solutions, and search order itself changes according to success and failure with a given search order in problems of the same type. Cyert and March also point out that the code or language in which an organization frames its problems changes through experience.

Simon views the organization as composed of boundedly rational agents and Cyert and March refer to the firm as an adaptively rational system. Much of the recent work on organizational learning is in this tradition. While we share the interest in how information is processed and decisions are rendered, the present paper explores what we can learn from a fully rational learning perspective. To the extent that equilibrium matters for the bounded rationality approach, one can perhaps characterize it as *learning an equilibrium,* whereas we investigate *learning in equilibrium.*

Organizational learning differs from individual learning because of decentralization of information and decisions. Because of decentralization of decisions, the learning activity of some agents may confound and frustrate the learning of others. This perspective on organizational learning is in line with the transaction cost view of the firm and in contrast with the neoclassical view of the firm as a single decision maker. Because of decentralized learning activities, organizational learning poses a complex coordination problem. Ideally, to learn efficiently, the organization may have to induce an intricate pattern of contingent decisions by its members. If decision making is diffuse, this may be next to impossible.

To formally represent the problem posed by decentralization, it is necessary but not sufficient to model the firm as a collection of independent strategic agents, i.e., a game. If we ignore conflicts of interest in the organization, the intricate decision pattern alluded to above may well be a Nash equilibrium of this game. The problem though is how to achieve the necessary coordination. In order to capture what differentiates organizational from individual learning, we need to formalize the constraints that decentralization imposes on the organization's ability to coordinate on an efficient learning equilibrium.

We model the decentralization constraints as lack of a common language. Following Crawford and Haller [1990], lack of a common language for a set of objects (e.g., actions in a game or decisions and agents in an organization) can be represented via symmetry restrictions on agents' beliefs about the strategies of other agents. One can modulate the symmetry constraints to represent various degrees of severity of the decentralization constraints. In addition, the history of repeated interaction in the organization may help in (partially) removing symmetry constraints. Thus part of the learning process consists of the construction of a common language to label decisions and agents in the organization.

We study the role of decentralization constraints for an organization that gathers information about which action profiles are optimal. The generation and processing of information is central to organizational activity. Prescott and Visscher [1980] go as far as saying that "The manner in which information is accumulated in the firm offers an explanation for the firm's existence." In Prescott and Visscher's setting a central agency in the organization accumulates the information. In our setting information entirely resides with the individual agents in the organization. Thus we are considering problems facing the "several-person firm" described by Marschak [1960], where each decision maker "decides about different things and on the basis of different information." Marschak [1960] and Milgrom and Roberts [1990a,b] emphasize independent decision making and the role of complementarities for organizational form. Marschak emphasizes the possibility of managers "step[ping] on each other's toes" and notes that complementarities may play a greater role when successive operations are required. Noting that there may be multiple optima, he observes that "two or more timetables are often equally good, but some 'co-ordinator' has to choose one."

The need to newly discover optimal action profiles arises in a changing environment. If these changes are routine, the discovery process may itself follow a possibly complicated routine, as if it were coordinated by a central agency. If changes are novel, e.g., because they frequently involve different subsets of agents in the organization, new agents in the organization, or completely new sets of available action, it is less likely that complicated learning routines develop. We offer a model of individual learning in an organization where novelty is captured through symmetry. In a novel situation agents are likely to lack at least some of the common distinctions necessary to implement intricate asymmetric learning schemes. The case of extreme symmetry, without any common distinctions of actions and agents, can be viewed as a benchmark for the value of coordination and of corporate culture.

Organizations need to learn to overcome absence of a common language each time they encounter novel circumstances. Recurrent novelty provides a role for (corporate) culture to guide this learning process. Culture can be viewed as a partial language, with some of the symmetries removed. It can serve as a focal point, Schelling [1960], in unanticipated situations. This is in line with Kreps' [1990] discussion of corporate culture as an answer to coordination difficulties that arise due to novel circumstances, or "unforeseen contingencies." Both Crémer [1993] and Hermalin [2000] make the connection between corporate culture and a common language. Crémer defines culture as common language plus shared knowledge of facts plus shared knowledge of simple rules. Hermalin [2000] argues that both coordination issues and unforeseen contingencies justify a role for culture. He adopts a partition view of a partial language that is consistent with our and Crawford and Haller's [1990] approach of using (partial) symmetry to express (partial) lack of a common language. Corporate culture assists organizational learning by labeling agents and decisions and thereby providing a scaffold for the construction of intricate learning schemes.

Crawford and Haller's approach to modeling language constraints through symmetry has been further developed and used in a number of recent papers. Blume [2000] studies partial languages with structure akin to a grammar or culture that enhances the use of the language in coordination and learning. Rubinstein [1996, 2000] uses an alternative approach for studying structure in language through properties of binary relations. Kramarz [1996] extends CH's work to n-player games. Bhaskar [2000] applies CH's approach to a game with conflict, the infinitely repeated *Battle of the Sexes.* Alpern and Reyniers [2000] study attainable strategies, subject to additional Markov restrictions, in games with many players whose goal is to disperse themselves among a finite set of locations.

Symmetry and language construction have also been looked at in the experimental literature. Blume, DeJong, Kim and Sprinkle [1998, 2001] experimentally study the emergence of meaning for *a priori* meaningless messages in repeated sender-receiver games. Blume and Gneezy [2000] find evidence for attainability in cognitively simple games. Blume and Gneezy

[2001] show that the attainability idea can also be used in settings where players do not have common knowledge of the language they are using. They show that players form beliefs about each other's languages and use *cognitive forward induction,* i.e., signal their language if given the opportunity. Weber and Camerer [2001] use a language construction experiment to study conflicting organizational cultures in the laboratory. In the experiment, "firms" develop some common language and are then merged. As one might expect, performance deteriorates immediately after the merger as a result of the conflict cultures, or local idioms. It takes time for the languages to be harmonized again.

Language provides both the means for communicating within the organization and the shared concepts used by members to infer how others think within the organization. While in this paper we focus on language constraints as constraints on common concepts in organizations, the importance of similar constraints on communication in organizations has been noted as well. The philosopher Putnam [1973] formulates the "hypothesis of the universality of the division of linguistic labor." Arrow [1974] stresses the local nature of an organization's code and links differences in codes to multiple equilibria and history dependence. Nelson and Winter [1982] speak of the "internal language of communication in an organization."

Learning in environments with multiple independent decision makers has recently attracted attention across a wide array of disciplines and with a large number of different approaches. Computer scientists, e.g., Sen and Sekaran [1998], Huhns and Weiss [1998] (for a brief survey) and Wellman and Hu [1998], have studied multi-agent learning, often with adaptive routines, such as reinforcement learning. On the boundary of computer science and psychology, there is work on *collaborative learning,* Dillenbourg [1999], which is, for example, interested in the "co-construction of a common language" (culture) in environments characterized by action symmetry, shared goals, and low division of labor into sub-tasks. Within cognitive science, there has been interest in *distributed cognition,* e.g., Hutchins [1995] and Hollan, Hutchins and Kirsh [1999]. This work emphasizes that the relevant cognitive entity is not necessarily the individual and studies organizational learning from this perspective. *Organizational learning* is also addressed in the strategic management literature; for a discussion of that literature and its link with the economics literatures on learning and the firm see, for example, Boerner, Macher and Teece [2000]. Closely related to the computer science work on multi-agent learning, there is work on organizational learning in computational organization theory, e.g., Carley [1998] and Carley and Hill [2001]. One of the themes there is to view organizations as synthetic agents, with knowledge existing in their culture and structure. Another branch of organization theory has been concerned with how to optimally structure hierarchies that search for optimal production plans, e.g., Beckmann [1977] and Keren and Levhari [1983]. Radner [1993] under some conditions derives hierarchies as efficient structures for decentralized information processing in organizations.

The bulk of the work just described invokes some kind of bounded rationality of the agents and sometimes argues that the cognitive ability of an organization may transcend that of the agents comprising it. In contrast, the present paper assumes that agents are fully rational and obtains its predictions from the interplay of rational learning, limited observability, and symmetry constraints. Fully rational learning in strategic settings is also investigated in the literature on many-agent versions of the multi-armed bandit problem, Bolton and Harris [1999], games with unknown payoff distributions, Wiseman [2000], and the literature on informational herding, e.g., Banerjee [1992] and Bikchandani, Hirshleifer and Welch [1992].

References

Alpern, S. and D.J. Reyniers [2000], "Spatial Dispersion in a Dynamic Coordination Problem," Department of Mathematics, London School of Economics Working Paper.

Arrow, K.J. [1974], *The Limits of Organization,* New York: W.W. Norton & Company.

Banerjee, A. [1992], "A Simple Model of Herd Behavior," *Quarterly Journal of Economics,* **107**, 797-818.

Beckmann, M.J. [1977], "Management Production Function and the Theory of the Firm," *Journal of Economic Theory,* **14**, 1-18.

Bhaskar, V. [2000], "Egalitarianism and Efficiency in Repeated Symmetric Games," *Games and Economic Behavior,* **32**, 247-262.

Bikchandani, S., D. Hirshleifer and I. Welch [1992], "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades," *Journal of Political Economy,* **100**, 992-1026.

Blume, A. [2000], "Coordination and Learning with a Partial Language,"*Journal of Economic Theory,* **95**, 1-36.

Blume, A., D.V. DeJong, Y.-G. Kim and G.B. Sprinkle [1998], "Experimental Evidence on the Evolution of the Meaning of Messages in Sender-Receiver Games," *American Economic Review,* **88**, 1323-1340.

Blume, A., D.V. DeJong, Y.G. Kim and G.B. Sprinkle [2001], "Evolution of Communication with Partial Common Interest," *Games and Economic Behavior,* **37**, 79-120.

Blume, A. and U. Gneezy [2000], "An Experimental Investigation of Optimal Learning in Coordination Games," *Journal of Economic Theory,* **90**, 161-172.

Blume, A. and U. Gneezy [2001], "Cognitive Forward Induction and Coordination without Common Knowledge: Theory and Evidence," University of Pittsburgh and Technion Working Paper.

Boerner, C., J.T. Macher, and D.J. Teece [2000], "A Review and Assessment of Organizational Learning in Economic Theories," in M. Dierkes, A. Berthoin-Antal, J. Child and I. Nonaka (eds), *Handbook of Organizational Learning and Knowledge,* New York: Oxford University Press.

Bolton, P. and C. Harris [1999], "Strategic Experimentation," *Econometrica,* **67**, 349-374.

Carley, K.M. [1998], "Organizational Adaptation," *Annals of Operations Research,* **75**, 25-47.

Carley, K.M. and V. Hill [2001], "Structural Change and Learning within Organizations," forthcoming in A. Lomi, *Dynamics of Organizational Societies: Models, Theories and Methods,* Cambridge: MIT Press.

Crawford, V. and H. Haller [1990], "Learning how to Cooperate: Optimal Play in Repeated Coordination Games," *Econometrica,* **58**, 581-596.

Crémer, J. [1993], "Corporate Culture and Shared Knowledge," *Industrial and Corporate Change,* **2**, 1993.

Cyert, R.M. and J.G. March [1963], *A Behavioral Theory of the Firm,* Englewood Cliffs: Prentice Hall.

Day H.D. and E.H. Tinney [1968], "How to Co-operate in Business without Really Trying: A Learning Model of Decentralized Decision Making," *Journal of Political Economy,* **76**, 583-600.

Dillenbourg, P. [1999], "What do you Mean by Collaborative Learning," in P. Dillenbourg (ed.) *Collaborative Learning: Cognitive and Computational Approaches,* Oxford: Elsevier Science.

Fudenberg, D. and D. Levine [1983], "Subgame-Perfect Equilibria of Finite and Infinite Horizon Games," *Journal of Economic Theory,* **31**, 251-268.

Fudenberg, D., D. Levine, and E. Maskin [1994], "The Folk Theorem with Imperfect Public Information," *Econometrica,* **62**, 997-1040.

Hermalin, B.E. [2000], "Economics and Corporate Culture," University of California - Berkeley Working Paper.

Hollan, J.D., E. Hutchins, and D. Kirsh [1999] "Distributed Cognition: A New Foundation for Human-Computer Interaction Research," University of California-San Diego Working Paper.

Huhns, M. and G. Weiss [1998], "Guest Editorial," *Machine Learning,* **33** 123-128.

Hutchins, E. [1995], *Cognition in the Wild,* Cambridge, MA: MIT Press.

Keren, M., and D. Levhari [1983], "The Internal Organization of the Firm and the Shape of Average Costs," *The Bell Journal of Economics,* **14**, 474-486.

Kramarz, F. [1996], "Dynamical Focal Points in N-Person Coordination Games," *Theory and Decision,* **40**, 277-313.

Kreps, D. [1990], "Corporate Culture and Economic Theory," in *Perspectives on Positive Political Economy,* eds. J. Alt and K. Shepsle, Cambridge: Cambridge University Press.

Marschak, J. [1960], "Theory of an Efficient Several-Person Firm," *American Economic Review,* **50**, 541-548.

Milgrom, P. and J. Roberts [1990a], "The Economics of Modern Manufacturing: Technology, Strategy and Organization," *American Economic Review,* **80**, 511-528.

Milgrom, P. and J. Roberts [1990b], "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, **58**, 1255-1277.

Munkres, J.R. [1975], *Topology: A First Course,* Englewood Cliffs, NJ: Prentice-Hall.

Nelson R.R. and S.G. Winter [1982], *An Evolutionary Theory of Economic Change,* Cambridge, MA: Harvard University Press.

North, D.C. [1994], "Economic Performance Through Time," *American Economic Review,* **84**, 359-368.

Obara, I. [2000], "Private Strategy and Efficiency: Repeated Partnership Games Revisited," University of Pennsylvania Working Paper.

Prescott, E.C. and M. Visscher [1980], "Organization Capital," *The Journal of Political Economy,* **88**, 446-461.

Putnam, H. [1973], "Meaning and Reference," reprinted from *The Journal of Philosophy* in E.P. Martinich (ed.) [1996], *The Philosophy of Language,* New York: Oxford University Press.

Radner, R. [1986], "Repeated Partnership Games with Imperfect Monitoring and no Discounting," *Review of Economic Studies,* **53**, 43-58.

Radner, R. [1993], "The Organization of Decentralized Information Processing," *Econometrica,* **61**, 1109-1146.

Radner, R., R. Myerson and E. Maskin, [1986], "An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," *Review of Economic Studies,* **53**, 59-70.

Rubinstein, A. [1996], "Why are Certain Properties of Binary Relations Relatively More Common in Natural Language?" *Econometrica,* **64**, 343-355.

Rubinstein, A. [2000], *Economics and Language,* Cambridge, UK: Cambridge University Press.

Schelling, T. [1960], *The Strategy of Conflict,* Cambridge, MA: Harvard University Press.

SEN, S. AND M. SEKARAN [1998], "Individual Learning of Coordination Knowledge," Working Paper, Department of Mathematical and Computer Science, University of Tulsa.

SIMON, H.A. [1947], *Administrative Behavior; A Study of Decision Making Processes in Administrative Organization; With a Foreword by Chester I. Barnard*, New York: Macmillan.

WEBER, R.A. AND C.F. CAMERER [2001], "Cultural Conflict and Merger Failure: An Experimental Approach," CMU and CalTech Working paper.

WELLMAN, M.P. AND J. HU [1998], "Conjectural Equilibrium in Multiagent Learning," *Machine Learning,* **33**, 179-200.

WISEMAN, T. [2000], "A Folk Theorem for Games with Unknown Payoff Distributions," Northwestern University Working Paper.