Paper to be presented at the

DRUID Society Conference 2014, CBS, Copenhagen, June 16-18

A Decentralized Learning Equilibrium

Andreas Blume
University of Arizona
Economics
ablume@email.arizona.edu

April M Franco
University of Toronto
Management
April.Franco@Rotman.UToronto.Ca

Abstract

Blume and Franco (2007) considers the search-for-success game where members of an organization search over various action-profi les for a success profi le. Each member has the same number of actions and common interests, but there is no communication between members before or after each round. Following Crawford and Haller (1990), we focused on attainable strategies, which are player-symmetric and action-symmetric. Because of the common interests aspect, players would like to search over as many di fferent action-profi les as quickly as possible, while the lack of communication prevents coordination and increases the probability that a particular action-profi le will be revisited before the success profi le is found. Here, we focus on the case of two players, each with two actions and no discounting. We consider the zero discount case, because it highlights the informational links across periods and is the only way to explicitly calculate an attainable equilibrium without placing arbitrary restrictions on the time horizon. The central result is that the optimal attainable strategy is one in which players randomize equally over both actions in odd periods and switch with probability one in even periods. In Blume and Franco, we show that an optimal strategy is also an equilibrium strategy which implies that the outcome is an equilibrium with endogenous cycles.

A Decentralized Learning Equilibrium*

Abstract

Blume and Franco (2007) considers the search-for-success game where members of an organization search over various action-profiles for a success profile. Each member has the same number of actions and common interests, but there is no communication between members before or after each round. Following Crawford and Haller (1990), we focused on attainable strategies, which are player-symmetric and action-symmetric. Because of the common interests aspect, players would like to search over as many different action-profiles as quickly as possible, while the lack of communication prevents coordination and increases the probability that a particular action-profile will be revisited before the success profile is found. Here, we focus on the case of two players, each with two actions and no discounting. We consider the zero discount case, because it highlights the informational links across periods and is the only way to explicitly calculate an attainable equilibrium without placing arbitrary restrictions on the time horizon. The central result is that the optimal attainable strategy is one in which players randomize equally over both actions in odd periods and switch with probability one in even periods. In Blume and Franco, we show that an optimal strategy is also an equilibrium strategy which implies that the outcome is an equilibrium with endogeneous cycles.

1   Introduction

Blume and Franco (2007) studied decentralized learning in organizations by us-
ing attainability constraints on agents' strategies in search-for-success games.
Attainability, first used in Crawford and Haller (1990), requires that strategies
satisfy both symmetry constraints on players and actions. Attainability helps to
capture the idea that members of an organization can be made interchangeable
and identical via corporate culture, as discussed in Krep (1986). In particular,
we are interested in understanding why we may observe recurrent change in
organizational behavior even though the economic envirnoment is unchanged.
We consider the case where members of the organization do not value the fu-
ture, even though the firm is infinitely lived. While the organization's discount
factor may be smaller than that of its members, the members' discount value
determines the strategies they take. In the optimal attainable Nash equilib-
rium, members optimize by taking some actions deterministically at times and
randomizing over actions at others. This stands in stark constrast with the
case where members have lexicographically prefer early success. In this case,
there is no Nash equilibrium.

For search-for-success games that are repeated over a long time horizon, the
analysis becomes increasingly intractable. In fact, in Blume and Franco, the
optimal attainable strategies were partially characterized in the infinite horizon
game for all possible discount factors; for an attainable strategy to be optimal
after any fixed time, agents must occasionally randomize over actions while at
the same time mixing probabilities cannot be uniformly bounded away from
zero. In addition, they show that optimal attainable strategies are equivalent
to attainable Nash equilibrium strategies for expected present discounted value
maximizers Here, we return to this issue of characterizing the optimal attainable
strategies and show that when the future is valued at zero and each agents
has only two actions, in equilibrium players alternate between deterministic
switching and randomizing. This result holds for small discount factors as well.
The case of small discount factors is interesting because, unlike in ordinary
repeated games, there are significant intertemporal links no matter how small is
the discount factor. The analysis of the limit lays bare the role of information
in determining play.

Next, we focus on the effect of changing to lexicographic preferences, where
agents have preference for early success and eliminating the discount factor.
The payoff link across periods slightly strengthened. While asymmetric strate-
gies are always preferred regardless of the preference structure because of the
intertemporal information links, they are not available under attainability. Un-
der lexicographic preferences, we find that there are no equilibrium strategies
that satisfy attainability. While in certain environments using lexicographic
preferences helps to find an economic solution, these preferences have the effect
of breaking link between the equilibrium strategies and the optimal ones. In
fact, the result is different from slightly incrementing the discount factor at zero.

In the following section, the search-for-success game is reviewed, along with
the attainability restrictions. In the following section, the two cases are con-

sidered; the infinite case where the future doesn't matter and the lexicographic case where early success is preferred. In the concluding section, the results are discussed and future avenues for research are considered.

2 The Game and Restrictions on Strategies

We use the repeated search-for-success game developed in Blume and Franco (forthcoming) but focus on the case with 2 players each of whom have 2 actions. Let X be the set of all possible assignments of successes to action profiles; the remaining profiles are failures. Before the first round, nature selects $\theta \in X$ for the game which determines which 1 profile is a success; this is a once-and-for-all selection for that game. Each player, $i \in \{1, 2\}$, has actions a_{ij}, $j = \{a, b\}$ and identical payoffs from each action combination $a = (a_{1j_1}, a_{2j_2})$. All players receive a payoff of 1 at a success profile and 0 when the action profile is a failure. Agents in the organization know the number of success profiles, but not their locations. All action profiles are ex ante equally likely to be success profiles.

The stage game is repeated in periods $t = \infty$, until a success is played once; the location of success profiles remains constant in each round of the game.[1] The repeated game is denoted as Γ. An agent can only observe her own actions and whether a success profile has been reached and can not observe the actions of the other player. As a result, players face a trade-off; by switching to a new action in the second period, they guarantee a new action-profile is visited in the current period, but lower the probability that a new action-profile will be visited in the following period. Players maximize the expected present discounted value of future payoffs with $0 \leq \delta < 1$. First we will consider preferences over strategy profiles that have a von Neumann-Morgenstern utility function representation when players are extremely impatient and then those with lexicographic preferences for early success while eliminating the discount factor.

Players' strategies can depend only on their private histories $h_i^t \in H_i^t$, where $H_i^t := A_i^t$ for all $t \geq 1$ where A_i^t is the set of all possible sequences of actions with t length and H_i^0 is a singleton set containing the null history h_i^0, since players can only observe success or failure in addition to their own actions. Since the game ends when a success profile is uncovered, the players need only keep track of their actions. The entire history ending in period t is denoted by $h^t := (h_1^t, ..., h_n^t)$ and an infinite history by h. Let H stand for the set of all infinite histories.

For a player i, a sequence of functions

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

is a behavior strategy. The set of probability distributions over A_i is denoted by $\Delta(A_i)$. The probability that player i's strategy assigns to his action a after

---

[1]This assumption sharpens the statement of results. Most of our results hold for a version of more standard repeated games that continue until the time horizon is reached regardless of the number of successes.

any history $h_i^{t-1} \in A_i^{t-1}$, and $a \in A_i$, is $f_i^t(h_i^{t-1})(a)$.

In order for the players to do well in this game, they must find a success quickly. As a result, players want to minimize the possibility of revisiting action profiles. In the case where of an organization that is able to coordinate the actions taken by its members, this would occur when new action profiles were taken in each period until the success profile was uncovered. In the case where the two member organization can not coordinate, the search process is complicated by the fact that each agent is unaware of what action the other member took in each period and each member took the incorrect 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, any two players i and i' will use identical strategies, up to a bijection between their action spaces, A_i and $A_{i'}$, and each player i's strategy is defined only up to a permutation of A_i. Formally, these two conditions amount to

(1)$\forall i, i' \in I$, $\exists$ a bijection $\hat{\rho}: A_i \to A_{i'}$ 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 $\tilde{\rho}: A_i \to A_i$.

The first condition is player symmetry and requires that players with identical histories (where actions are named by the periods where they were first taken) behave identically. The second condition is action symmetry and requires that players put equal weight on all actions that have not been taken previously; only actions that have been taken previously may be treated differently by a player. Strategy profiles that satisfy these restrictions are *attainable*. *Optimal attainable strategy profiles* are strategy profiles that maximize players' payoffs among the set of attainable strategies, while individual strategies that satisfy this are *optimal attainable strategies. Equilibrium attainable strategy profiles* are attainable strategy profiles that are Nash equilibria. Whether the strategy is optimal or equilibrium, the symmetry restrictions apply only on the solution path; when checking for equilibrium, we impose no restrictions on deviations. If individual strategies satisfy the second condition, they are attainable. As a result, we will limit our discussion to individual strategies rather than entire profiles.

Taking into account that a success ends the game, any behavior strategy profile f induces a probability $\pi_t(f)$ of a success in period t. It suffices to write this probability as $\pi_t(f^1, \ldots, f^t)$, that is, as a function of behavior only up to and including period t. In the discounting case, each player's payoff from profile f is

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

Next, we turn to the case where agents do not value the future at all.

3   The Case of Expected Present Value Maximizers

For small discount factors, we can approximately characterize the solution for any horizon, unlike the case studied in Blume and Franco where for large time horizons the analysis of the general version of the search-for-success game becomes intractable. 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. Let σ specify a switching probability $\sigma(h^t)$ conditional on each individual history h^t ending in period $t-1$. In addition, we assume that 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 1. *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, in every optimal attainable strategy σ the switching probability is one in even periods, and in odd periods both the switching probability conditional on even histories and the switching probability conditional on odd histories equal one-half.*

Proof. We show that any attainable strategy σ whose switching probabilities conditional on odd and even histories equal one in even periods and whose switching probability in odd periods equals one-half is an optimal attainable strategy. We proceed by induction. We first show that switching with probability one is uniquely optimal in period two and that conditional on having switched with probability one in period two, it is uniquely optimal to switch with probability one-half in period three. 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 period, randomization is constrained to assigning probability one-half to either action. If the first 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 2. 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 two. 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 3, the goal in period 3 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 3 by σ_3, the probability of success in that period is $\frac{1}{2}\sigma_3(1-\sigma_3)$, 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 odd. Since period t is odd, by assumption both players switched with probability one in the preceding period. Since we are in period five 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 odd and the other player's history is even. 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

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

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

This function attains its maximum whenever

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

It remains to show that if t is even 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 odd (even), 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 even and σ has been followed until period $t-1$, then the success probability in period t is

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

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

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

and similarly

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

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 $\sigma(h) = 1 \ \forall h \in H^{ot}$. □

Note that by extending the number of actions for each player increases the complexity of the problem. Since $\delta = 0$, clearly the agents will choose new actions in the first m periods. In the $m + 1$ period, they will randomize across all actions. However, after that, the problem becomes much more difficult and is not easily characterized. A final note regarding the infinite case with zero discounting, no public optimal attainable strategy exists just as in the three period case considered in Blume and Franco.

Next we show that for small δ, the optimal attainable strategy converges to the same one as when $\delta = 0$. In other words, agents alternate between deterministic behavior and randomizing. This helps to show the role of the informational links in this problem.

Proposition 2. *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:

$$\Pi(\sigma, \delta) = \frac{1}{4} + \sum_{t=2}^{\infty} \delta^t - 1\pi_t(\sigma_1, \ldots, \sigma_t),$$

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 even $\tau < t$, and $\sigma(H^{\alpha\tau}; 0) = \frac{1}{2}$ for $\alpha = e, o$ and all odd $\tau < t$, 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

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

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 $\sigma(\delta_{n_k})$. $\qquad \square$

Proposition 2 implies that for discount factors close to 0, the optimal strategy is the same as the one prescribed in Proposition 1. In Blume and Franco, we showed that an optimal attainable strategy was also an equilibrium attainable strategy; if the optimal attainable strategy was not an equilibrium attainable strategy, then there would be a profitable deviation which would lead to a contradiction. Based on that result, we have the following corollary.

Corollary 3. *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.*

9

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

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

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 $\pi_{t^*}^i > \hat{\pi}_{t^*}^i$.

Proposition 4. *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 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. $\square$

4   Conclusion

We characterize the optimal attainable strategies in the infinite horizon search-for-success game in the extreme impatience case. We find that, unlike the case in Blume and Franco, the optimal attainable strategy can be characterized for $\delta = 0$ and show what the optimal attainable strategy converges to as the limit approaches 0. In the case with zero discounting, the optimal attainable equilibrium with two actions, two players and one success profile has an interesting simple structure: agents switch between not repeating an action every other period and taking either action with equal probability. Even though the economic envirnoment is unchanged, the agents either switch actions with probability one or one-half in response to failure. The optimal strategy includes deterministic behavior but is not always deterministic which was Blume and Franco showed

to be a characteristic of the optimal attainable strategy. All solutions for sufficiently small δ approximate this $\delta = 0$ solution. Finally, in this environment, if we remove the discounting of the future and give agents lexicographic preferences for early success, the optimal attainable strategy remains the same, but no attainable Nash equilibrium exists.

References

[1] BLUME, ANDREAS, AND APRIL MITCHELL FRANCO [2007], "Decentralized Learning from Failure," *Journal of Economic Theory* **1**, 504-523.

[2] CRAWFORD, V. AND H. HALLER [1990], "Learning How to Cooperate: Optimal Play in Repeated Coordination Games," *Econometrica,* **58**, 581-596.

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