Rational Multi-Agent Search

Andreas Blume
Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260

April Franco
Department of Economics
University of Iowa
Iowa City, IA 52242

Paul Heidhues
Department of Economics
University of Bonn
Adenauerallee 24-42
Bonn
and CEPR

Abstract

We study games in which players search for an optimal action profile. All action profiles are either a success, with a payoff of one, or a failure, with a payoff of zero. Players do not know the location of success profiles. Instead each player is privately informed about the marginal distribution of success profiles over his actions. We characterize optimal joint search strategies.

---

* This is a preliminary draft that is currently under construction. We very much appreciate comments.
November 4, 2005

1 Introduction

This paper investigates *rational multi-agent search for an optimal profile of actions through trial and error.* The focus is on a benchmark case that already poses considerable difficulties: agents are assumed to have common interest and not to communicate.

The most closely related work in the literature is on repeated coordination games, where payoffs are known and participants have repeated opportunities to coordinate on an optimal combination of actions (e.g. an industry standard), and on rendezvous search problems, where the participants' goal is to find each other in the shortest time possible on some geographical domain. The innovation in the present paper is to consider strategic situations in which (1) the payoffs from each action are not known, but have to be discovered, (2) agents search not for each other but for an optimal joint choice of actions, and (3) agents have private information about which of their actions are likely to contribute to a success.

Situations like this abound in practice. Consider for example the coordination problem between the marketing department and the production department of a firm when there is uncertainty about how consumers respond to a given mix of product characteristics and marketing strategy, or the coordination problem between separate government agencies facing a common issue (poverty, environmental degradation, a terrorist threat) when there is uncertainty about the optimal policy mix.

Frequently in these settings search strategies will be imperfectly coordinated and information will be imperfectly shared. A prominent recent example is the lack of information sharing between law enforcement agencies that has at least partly been blamed for the failure to prevent the September 11 attacks. While improved communication is often helpful, there may be cost and institutional reasons that prevent the establishment of the appropriate infrastructure for large-scale information exchange. One obvious example is that of an auditing firm serving multiple clients; the auditing firm may need to erect "Chinese walls" to safeguard the proprietary information of its clients. For these reasons, this paper investigates the benchmark case where agents can neither share their private information nor explicitly coordinate their joint search.

The broader goal of this research is to better understand such issues as: the simultaneous search for optimal policies by multiple government agencies; the problem of how to best counteract firms who tacitly search for illegal collusive arrangements in markets; the problem of how to disrupt simultaneous search efforts by members of clandestine organizations; and, the effect of barriers to information exchange, so-called "Chinese Walls," on simultaneous search efforts by separate departments in an organization (as practiced e.g. in financial, legal and auditing firms).

2 Examples and Applications

In this section, we give additional examples and describe a variety of applications that help illustrate the class of problems we are interested in.

A stylized example of the situations we have in mind is that of a group of individuals wanting to open a safe, when each has access to a separate dial, there is one unknown combination of settings of the dials that will open the safe, the individuals are visually separated, and they have to avoid communication in order not to be discovered. Communication would allow them to coordinate their search and perhaps to share private information of which settings are more likely to be the correct combinations. Absent communication, it may be difficult to properly reflect private information in the search order and to avoid combinations that have been tried before.

Another example is that of two parties trying to coordinate on a radio frequency. In that case, since communication is the objective, it cannot help in coordinating the search. We are concerned with the situation where the parties may have private information about, but do not know, which frequencies are available (some of the frequencies may be jammed). If the available frequencies were commonly known, this would be a version of the *telephone problem* that was proposed by Steve Alpern (see Alpern [1976]). In the telephone problem, there is an equal number of telephones in two rooms. They are pairwise connected. In each period a person in each room picks up the receiver on one of the phones. The goal is to identify a working connection according to a payoff criterion that favors early discovery, e.g. minimum expected time or expected present discounted value.

An applied health economics example is that of multiple physicians simultaneously treating the same patient. Here communication may be prevented by simple lack of time, inadequate technology for file sharing, or privacy concerns. The physicians may be uncertain about which treatment combination is likely to be effective. This problem is similar to that of multiple consultants advising the same company on different dimensions and who may be uncertain of which (multi-dimensional) course of action is likely to succeed. Despite a common interest in the success of the company, they may not want to or are unable to share proprietary information. Similarly, an organization with multiple departments searching for an optimal course of action may have rules preventing exchange of information between departments (so-called "Chinese walls") if such an informational exchange could involve a conflict of interest, e.g. an accounting firm operating both as an auditor of a firm and as a consultant for that firm, or an accounting firm working for multiple clients and handling those clients' proprietary information.

Situations in which searching agents have to be secretive fit naturally into our environment without communication. One example would be battlefield coordination in which different units have to operate under radio silence, another coordination between members of a clandestine organization. Economic versions of such examples arise naturally in collusion between firms, where communication is explicitly illegal. Imagine two firms choosing product characteristic who are afraid that characteristics that are close in the minds of consumers increase price competition. Firms may have beliefs about, but not know, which characteristics are perceived as close by consumers.

The common interest assumption is made because we want to understand information aggregation and processing, which turns out to be delicate enough even without confounding it with payoff differences. This approach is not unusual. A prominent example is provided by the literature on the Condorcet jury theorem and related issues of information aggregation, e.g. McLennan [1998]. This approach is also pursued in the literature on *rendezvous search* where two or more parties attempt to find each other in the least time in some search space subject to a constraint on their speed (this literature is surveyed in Alpern [2002] and Alpern and Gal [2003]).

Broadly, the problem that interests us is related to other models of rational learning of payoffs in games, e.g. Wiseman's [2005] work on repeated games with unknown payoff distributions and Gossner and Vieille's [2003] work on games in which players learn their payoffs from experience. Another prominent example is the work on social learning, e.g. Banerjee [1992], Bikchandani, Hirshleifer and Welch [1992], and the recent book by Chamley [2004]. The problem addressed in the present paper – joint search for a hidden success with private information about the success location – is novel. A direct antecedent is Blume and Franco [2005] who study joint search for a success profile under complete information and with a symmetry constraint on joint strategies, and Blume, Duffy and Franco [2004] who look at the same problem experimentally.

3 Setup

We consider games in which players have payoff uncertainty and have private information about the distribution of payoffs. Each player has a finite set of actions A_i with cardinality $J(i)$. Let $A := \times_{i=1}^{n} A_i$ denote the set of action profiles. There is a set of payoff functions U with typical element $u \in U$. Any pair of an action profile $a \in A$ and a payoff function $u \in U$ determines a payoff $u_i(a)$ for player i. Players face payoff uncertainty. In order to model this uncertainty, we introduce the set of distributions $\Delta(U)$ over the set of possible payoff functions U. In addition we want to consider players who have private information about the nature of the payoff uncertainty. For this purpose, we introduce another layer of uncertainty: There is a common knowledge distribution F over the set $\Delta(U)$ of possible distributions over payoff functions. Players choose actions in each of T periods, where T may be infinite. At the end of each period, each player learns only his own payoff for that period. Players evaluate payoffs in the T-period game according to the expected present discounted value of the sum of stage-game payoffs, with common discount factor $\delta \in (0,1)$. The timing is as follows: (1) Nature draws a distribution $D \in \Delta(U)$ from the distribution F. (2) Each player receives a signal $s_i(D) \in S_i$ that is a function of the realized distribution D. (3) A payoff profile u is drawn from the realized distribution D. (4) The T-period game starts.

Unless otherwise stated, we focus on the subclass of *independent search-for-success* games in which (1) U consists of all those payoff functions that give all players a payoff of one for exactly one action profile, a *success*, and otherwise a payoff of zero, a *failure*, (2) for each player i and all distributions D in $\Delta(U)$ player i's marginal distribution D_i is exchangeable, (3) each player's signal $\omega_i(D)$ is the marginal distribution of D over player i's actions, and (4) each D is the product of its marginals. Given exchangeability, it is without loss of generality to label each player i's actions by the probability rank of these actions according to the signal that he receives, i.e. given signal ω_i, action $a_{i1}(\omega_i)$ is player i's action with the highest probability of being part of a success profile, $a_{i2}(\omega_i)$ the action with the second-highest probability, and so on. Then, we can denote by ω_i the vector of marginal probabilities observed by individual i where the individual probabilities have been put in rank order, i.e. ω_{ij} is i's marginal probability that his action $a_{ij}(\omega_i)$ of jth rank is part of a success profile. Whenever there is no risk of confusion, we will suppress the dependence of player i's jth ranked action, and simply write a_{ij} for $a_{ij}(\omega_i)$.

4 Example: Two Players, Two Actions, Two Periods and Independent Uniform Signal Distributions

This section introduces the class of games of interest by way of a $2 \times 2 \times 2$-*independent-uniform-example*, i.e. with two players, two actions per player, two choice periods, signals that are independent across players, and a uniform signal distribution. The example serves to develop intuition that carries over to more general games with an arbitrary (finite) number of players and actions per player, and with an infinite time horizon. In the example, one can fully characterize equilibrium and optimal behavior and illustrate the difficulties arising in joint search more generally. Specifically, one can show that there are multiple Pareto-ranked equilibria, with the optimal equilibrium bounded away from *ex post* efficient search.

Let each of two players repeatedly choose between two actions. It is common knowledge between them that exactly one of the four action profiles is a "success," but not which one. The other profiles are failures. The common payoff from a success is (normalized to) one and the common payoff from a failure is zero. The game ends when a success has been found, or the time horizon has been reached. Players have a common discount factor δ, so that a success in period t is worth δ^t. Players do not observe each others' actions and therefore when making a choice in any period can condition only on the history of their own actions.

Suppose in addition that before the first period each player privately receives a signal that informs him of the probability with which each of his actions is part of a success profile. For simplicity, let signals be independent across players. Then, if player one receives the signal ω_{1j} that his action a_{1j} is part of a success profile and player two receives the signal ω_{2k} that his action a_{2k} is part of a success profile, the probability that the profile (a_{1j}, a_{2k}) is a success is $\omega_{1j} \cdot \omega_{2k}$. Simplify further by assuming that player one's (two's) signal ω_{11} (ω_{21}) that his first action is part of a success is uniformly distributed on the interval $[0, 1]$.

In this section, in order to simplify notation, for player one, denote the higher of his two signals by α, i.e. $\alpha := \omega_{11}$ Similarly, for player two, define $\beta := \omega_{21}$. α is the first order statistic of the uniform distribution on the one-dimensional unit simplex. Note that α and β are independently and uniformly distributed on the interval $[\frac{1}{2}, 1]$. In the sequel, when talking about player one's action, it will be convenient to refer to his α (or high-probability) action and his $1 - \alpha$ (or low probability) action, and similarly for player two.

Simplifying once again, consider the two-period game. In this game, players only have to decide which action to choose in the first period and whether or not to switch to a different action in the second period if the first-period choice fails to deliver a success.

It is immediately clear that the full-information solution (or *ex post*-efficient search), which a social planner with access to both players' private information would implement, is not an equilibrium in the game with private information. The social planner would prescribe the α-action to player one and the β-action to player two in the first period, and in the second period would prescribe the profile $(\alpha, (1 - \beta))$ if $\alpha(1 - \beta) > (1 - \alpha)\beta$, and the profile $((1 - \alpha), \beta)$ otherwise. The players themselves, who only have access to their own information, are unable to carry out these calculations and cannot decide which of the two players should switch actions and who should stick to his first-period action.

This raises a number of questions: What is the constrained planner's optimum, i.e. which strategy profile would a planner prescribe who does not have access to the players' private information? What are the Nash equilibria of the game? What is the relationship between the (constrained) planner's solution and the Nash equilibria of the game?

Since this is a common interest game, i.e. the payoff functions of the players coincide, there is a simple relation between optimality and equilibrium. An optimal strategy profile must be a Nash equilibrium (see Alpern [2002], Crawford and Haller [1990], and McLennan [1998]). A useful corollary of this observation is that if we have a Nash equilibrium and there is another strategy, with higher payoffs for both players, then either the latter strategy is an equilibrium or there exists an equilibrium with even higher payoffs for both players.

Two strategy profiles are easily seen to be equilibria. In one, player one takes his α action in both periods and player two takes his β action in the first and his $1-\beta$ action in the second period. In the second equilibrium, player two stays with his β action throughout and player one switches. In these equilibria only the first-period decision is sensitive to the players' information; the switching decision does not depend on the signal. One may wonder whether it would not be better to tie the switching probability to the signal as well. Intuitively, a player one with a strong signal, α close to one, should be less inclined to switch than a player with a weak signal, α close to one half. In order to investigate the possibility of such equilibria, we need to formally describe players' strategies.

A strategy for player i has three components: (1) $p^i(\alpha)$, the probability of taking the high-probability action in period 1 as a function of the signal; (2) $q^i_h(\alpha)$, the probability of taking the high-probability action in period 2 after having taken the high-probability action in period 1, as a function of the signal; and (3), $q^i_l(\alpha)$, the probability of taking the high-probability action in period 2 after having taken the low-probability action in period 1, as a function of the signal.

Fix a strategy for player 2. We are interested in the payoff of player one for anyone of his types, for any possible action sequence he may adopt, and for any possible strategy of player two. In writing down payoffs, we will use the fact that in equilibrium player two will never stick to his low-probability action in the second period after having used his low-probability action in the first period, i.e. $q^2_l(\beta) = 1$ for all $\beta \in [\frac{1}{2}, 1]$ in every equilibrium. Then type α of player 1 has the following payoff from taking the high-probability action in both periods:

$$\begin{aligned} \text{(1)} \qquad \mathrm{HH}(\alpha) &= \int_{\frac{1}{2}}^{1} \left[\alpha\beta + \alpha(1-\beta)\delta(1-q^2_h(\beta))\right] p^2(\beta)d\beta \\ &+ \int_{\frac{1}{2}}^{1} \left[\alpha\beta\delta + \alpha(1-\beta)\right](1-p^2(\beta))d\beta \end{aligned}$$

Player 1's payoff from taking the high-probability action in the first and the low-probability action in the second period, when his type is α, equals

$$
\text{(2)} \qquad \mathrm{HL}(\alpha) = \int_{\frac{1}{2}}^{1} \left[\alpha\beta + (1-\alpha)\beta\delta q_h^2(\beta) + (1-\alpha)(1-\beta)\delta(1-q_h^2(\beta))\right] p^2(\beta)d\beta + \int_{\frac{1}{2}}^{1} \left[\alpha(1-\beta) + (1-\alpha)\beta\delta\right](1-p^2(\beta))d\beta
$$

Player 1's payoff from taking the low-probability action in the first and the high-probability action in the second period, when his type is α, equals

$$
\text{(3)} \qquad \mathrm{LH}(\alpha) = \int_{\frac{1}{2}}^{1} \left[\alpha\beta\delta q_h^2(\beta) + \alpha(1-\beta)\delta(1-q_h^2(\beta)) + (1-\alpha)\beta\right] p^2(\beta)d\beta + \int_{\frac{1}{2}}^{1} \left[\alpha\beta\delta + (1-\alpha)(1-\beta)\right](1-p^2(\beta))d\beta
$$

The sequence of actions LL is strictly dominated for all $\alpha > \frac{1}{2}$.

Simple inspection yields the following observations about these three payoffs: All three payoffs are linear in α; $\mathrm{HH}(\cdot)$ is strictly increasing in α; $\mathrm{HH}(1) \geq \mathrm{LH}(1)$ (with equality only at $\delta = 1$); $\mathrm{HH}(1) \geq \mathrm{HL}(1)$; $\mathrm{HL}(\frac{1}{2}) \geq \mathrm{HH}(\frac{1}{2})$; and, $\mathrm{HL}(\frac{1}{2}) = \mathrm{LH}(\frac{1}{2})$.

This implies that in equilibrium player 1 (similarly for player 2) either plays HH for all α, or HL for all α, or LH for all α, or there exists a critical value c_1 such that he plays HL for $\alpha \leq c_1$ and HH for $\alpha > c_1$, or there exists a critical value c_1 such that he plays LH for $\alpha \leq c_1$ and HH for $\alpha > c_1$. In addition, against a player using only the action sequences HH and HL, the action sequence LH is never optimal, because in that case HL is a better response. This leaves only two possible types of equilibria:

1. *HL-equilibria* in which player i has a cutoff c_i such that he uses HL for α below this cutoff (and HH above the cutoff), and

2. *LH-equilibria* in which player i has a cutoff c_i such that he uses LH for α below this cutoff (and HH above the cutoff).

The cutoffs in any HL-equilibrium must satisfy the system of equations:

$$
\text{(4)} \qquad \int_{c_{3-i}}^{1} c_i\beta d\beta + \int_{\frac{1}{2}}^{c_{3-i}} \left[c_i\beta + c_i(1-\beta)\delta\right] d\beta = \int_{c_{3-i}}^{1} \left[c_i\beta + (1-c_i)\beta\delta\right] d\beta + \int_{\frac{1}{2}}^{c_{3-i}} \left[c_i\beta + (1-c_i)(1-\beta)\delta\right] d\beta \qquad i = 1, 2.
$$

This system has exactly three solutions in the relevant range of $c_i \in [\frac{1}{2}, 1]$ $i = 1, 2$. These are, $(c_1, c_2) = (.5, 1)$, $(c_1, c_2) = (1, .5)$, and $(c_1, c_2) \approx (0.760935, 0.760935)$.

The cutoffs in any LH-equilibrium must satisfy the system of equations:

$$
\begin{aligned}
(5) \qquad & \int_{c_{3-i}}^{1} c_i \beta d\beta + \int_{\frac{1}{2}}^{c_{3-i}} \left[c_i \beta \delta + c_i(1-\beta)\right] d\beta \\
= & \int_{c_{3-i}}^{1} \left[c_i \beta \delta + (1-c_i)\beta\right] d\beta + \int_{\frac{1}{2}}^{c_{3-i}} \left[c_i \beta \delta + (1-c_i)(1-\beta)\right] d\beta \qquad i = 1, 2.
\end{aligned}
$$

In this case the solutions of this system of equations in the relevant range of $c_i \in [\frac{1}{2}, 1]$ $i = 1, 2$, depend on δ. For $\delta = 1$, there are three solutions: $(c_1, c_2) = (.5, 1)$, $(c_1, c_2) = (1, .5)$, and a symmetric solution. Otherwise, there is a unique solution, which is symmetric, i.e. there is a common cutoff $c(\delta)$, which is a strictly increasing function of δ, with $c(0) = .5$.

The fact that there is a unique LH-equilibrium for each $\delta \in (0, 1)$ can be shown as follows: Substitute one of the equations in (5) into the other in order to eliminate the variable c_2. This leaves, for any δ, a fifth-order polynomial $\phi(\cdot)$ in the variable c_1, which has at most five real roots and at most four local extrema. The relevant roots are the ones in the interval $[\frac{1}{2}, 1]$. The function $\phi(\cdot)$ satisfies (1) $\phi(\frac{1}{2}) < 0 \;\; \forall \delta \in (0, 1)$ and (2) $\phi(1) > 0 \;\; \forall \delta \in (0, 1)$. Differentiating $\phi(\cdot)$ gives a fourth-order polynomial $\psi(\cdot)$ which satisfies for all $\delta \in (0, 1)$ the following conditions: (1) $\psi(c_1) > 0$ for sufficiently small c_1 $(c_1 \to -\infty)$, (2) $\psi(.2) < 0$, (3) $\psi(.7) > 0$, (4) $\psi(1) < 0$, and (5) $\psi(c_1) > 0$ for sufficiently large c_1 $(c_1 \to \infty)$.

As local extrema of ϕ correspond to the roots of ψ, one local extremum (a maximum) of ϕ must be in the interval $(-\infty, .2)$, another (a minimum) in the interval $(.2, .7)$, another (a maximum) in the interval $(.7, 1)$, and another (a minimum) in the interval $(1, \infty)$. Using the fact that $\phi(\frac{1}{2}) < 0$ and $\phi(1) > 0$, and that ϕ has only one local minimum in the interval $(.2, .7)$ and only one local maximum in the interval $(.7, 1)$, the equation $\phi(c_1) = 0$ has a unique solution for $c_1 \in [\frac{1}{2}, 1]$. Thus there exists only one equilibrium, which must be symmetric as c_2 has to satisfy the same equation. We can summarize our discussion by the following observation:

Proposition 1 *For any $\delta \in (0, 1)$, the two-player two-action two-period game with signals that are independently and uniformly distributed has exactly four Nash equilibria: One symmetric HL-equilibrium with common cutoff $c \approx 0.760935$, two asymmetric HL-equilibria with cutoffs $(c_1, c_2) = (.5, 1)$ and $(c_1, c_2) = (1, .5)$, respectively, and one symmetric LH-equilibrium with common cutoff $c(\delta)$ that is strictly increasing in δ.*

An LH-equilibrium cannot be optimal: To see this, simply change both players' strategies to HL-strategies, without changing the cutoff. Under the original strategies, there are three possible events, each arising with strictly positive probability: Both players follow an HH-sequence; both follow an LH sequence; and, one follows an LH-sequence while the other follows an HH sequence. Clearly LH is not optimal against HH and therefore in this instance the new strategy yields a strict improvement. Also, both players following HL rather than LH yields a strict improvement. Thus in two events there is a strict payoff improvement, in the remaining event payoffs are unaffected, and all three events have strictly positive probability.

It is not immediately clear whether to prefer the symmetric HL-equilibrium or the asymmetric HL-equilibria. In either, there is positive probability that profiles are searched in the wrong order. The symmetric equilibrium makes the second-period switching probability sensitive to a player's signal, which seems sensible. At the same time, it introduces an additional possible source of inefficiency. Players may not succeed in the first round despite having signals so strong that they do not switch in the second round. In that case, they inefficiently search only one of the available profiles.

It would be a straightforward matter to calculate and compare payoffs from symmetric and asymmetric equilibria directly. We will follow a different route in order to introduce some methodological ideas that may prove useful more generally. Start with the asymmetric HL-equilibrium in which $c_1 = \frac{1}{2}$ and $c_2 = 1$. Consider the (informationally constrained) social planner who raises c_1 from $\frac{1}{2}$ and lowers c_2 from 1 by the same amount γ. His second-period payoff (note that the first-period payoff is not affected by γ) as a function of γ is proportional to

$$\pi(\gamma) = \int_{\frac{1}{2}}^{1-\gamma} \int_{\frac{1}{2}}^{\frac{1}{2}+\gamma} (1-\alpha)(1-\beta) d\alpha d\beta + \int_{1-\gamma}^{1} \int_{\frac{1}{2}}^{\frac{1}{2}+\gamma} (1-\alpha)\beta d\alpha d\beta + \int_{\frac{1}{2}}^{1-\gamma} \int_{\frac{1}{2}+\gamma}^{1} \alpha(1-\beta) d\alpha d\beta.$$

It is straightforward to check that $\left.\frac{\partial \pi(\gamma)}{\partial \gamma}\right|_{\gamma=0} = 0$ and $\left.\frac{\partial^2 \pi(\gamma)}{\partial \gamma^2}\right|_{\gamma=0} > 0$. Hence, the social planer can improve on the two asymmetric equilibria. Recall that for any arbitrary strategy profile σ, either σ is an equilibrium or there exists an equilibrium σ^* with $u_i(\sigma^*) > u_i(\sigma)$ for $i = 1, 2$. Thus the pair of cutoff strategy profiles with cutoffs $c_1 = 1/2 + \gamma$ and $c_2 = 1 - \gamma$ with an appropriately small value of γ either is an equilibrium or there exists an equilibrium that strictly dominates it. Therefore, we have the following observation:

Proposition 2 *For any $\delta \in (0, 1)$, in the two-player two-action two-period game with signals that are independently and uniformly distributed, the symmetric HL-equilibrium is the optimal equilibrium and at the same time the optimal strategy that an informationally constrained social planner would implement.*

4.1 The 2×2 Uniform Example with an Infinite Horizon

In the infinite-horizon game there are equilibria in which players conduct a joint exhaustive search without repetition. This search is guaranteed to be successful in no more than four periods.

We will show for a range of values of the discount factor that (1) these exhaustive-search equilibria are not optimal for a social planner who has to respect the same informational constraints as the players, (2) there exists an equilibrium in the infinite horizon-game with higher expected payoff than from exhaustive search and (3) that a payoff-maximizing equilibrium will lead to repetition of action profiles with positive probability.

There are exactly two exhaustive-search equilibria. They simply exchange the roles of the two player and thus have identical payoffs. For concreteness, we argue with reference to the exhaustive search equilibrium in which only player switches in period 2, both players switch in period 3 and only player 2 switches in period 4, i.e. in terms of their signals, players move through cells in the following order: $(\alpha, \beta), (\alpha, 1-\beta), (1-\alpha, \beta), (1-\alpha, 1-\beta)$. We proceed by showing that the following strategy leads to a higher expected payoff than exhaustive search for some values of δ: *Pick the likelihood-maximizing action in the first period. Use a symmetric cutoff rule with cutoff value c in the second period.*[1] *Revert to the exhaustive-search equilibrium from the third period on (i.e. if there is no success before period three, resume exhaustive search in the cell $(\alpha, 1-\beta)$).* Refer to this strategy as modified exhaustive search.

The payoff from modified exhaustive search equals:

[1] An alternative approach would be to adopt the methodology from the previous subsection, i.e. to look at small departures from exhaustive search equilibria, slightly lowering the switching player's cutoff value from 1 and slightly raising the other player's cutoff value from 1/2

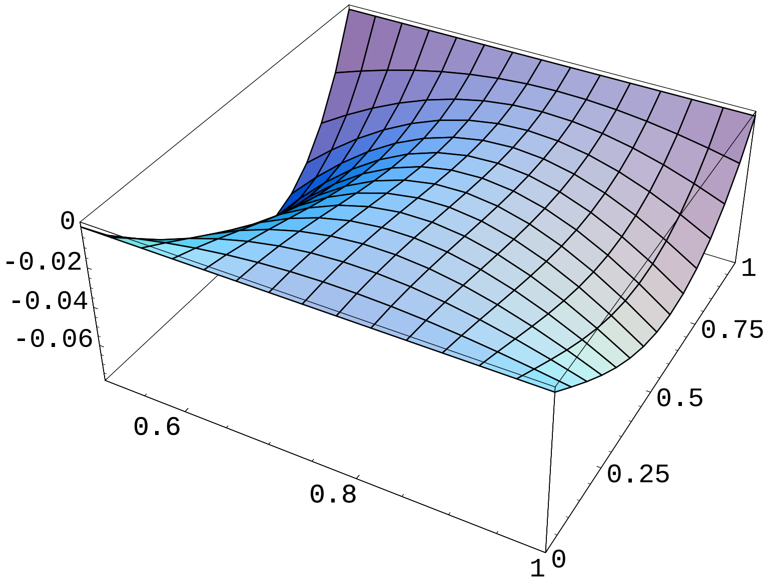
$$
\begin{aligned}
& 4\int_{\frac{1}{2}}^{c}\int_{\frac{1}{2}}^{c}[\alpha\beta+\delta(1-\alpha)(1-\beta)+\delta^2\alpha(1-\beta)+\delta^3(1-\alpha)\beta]d\alpha d\beta \\
+ \quad & 4\int_{c}^{1}\int_{\frac{1}{2}}^{c}[\alpha\beta+\delta(1-\alpha)\beta+\delta^2\alpha(1-\beta)+\delta^4(1-\alpha)(1-\beta)]d\alpha d\beta \\
+ \quad & 4\int_{c}^{1}\int_{c}^{1}[\alpha\beta+\delta^2\alpha(1-\beta)+\delta^3(1-\alpha)\beta+\delta^4(1-\alpha)(1-\beta)]d\alpha d\beta \\
+ \quad & 4\int_{\frac{1}{2}}^{c}\int_{c}^{1}[\alpha\beta+\delta\alpha(1-\beta)+\delta^3(1-\alpha)\beta+\delta^4(1-\alpha)(1-\beta)]d\alpha d\beta \\
= \quad & (c^2-\frac{1}{4})^2+\delta((1-c)^2-\frac{1}{4})^2-\delta^2(c^2-\frac{1}{4})((1-c)^2-\frac{1}{4})-\delta^3((1-c)^2-\frac{1}{4})(c^2-\frac{1}{4}) \\
= \quad & (1-c^2)[(c^2-\frac{1}{4})-\delta((1-c)^2-\frac{1}{4})]+(1-c)^2[\delta^2(c^2-\frac{1}{4})-\delta^4((1-c)^2-\frac{1}{4})] \\
= \quad & (1-c^2)[(c^2-\frac{1}{4})-\delta((1-c)^2-\frac{1}{4})]+(1-c)^2[\delta^3(c^2-\frac{1}{4})-\delta^4((1-c)^2-\frac{1}{4})] \\
= \quad & (1-c^2)^2+\delta^2(1-c^2)(1-c)^2+\delta^3(1-c)^2(1-c^2)+\delta^4(1-c)^4
\end{aligned}
$$

For example, the second line of the above expression is the contribution to expected payoffs from the event that player 1 receives a signal $\alpha < c$ and player 2 receives a signal $\beta > c$. In that event, a success will be found in the first period with probability $\alpha\beta$; in the second period player 1 switches (because of his relatively low signal) while player 2 stays put leading to a success probability of $(1-\alpha)\beta$; in the third period exhaustive search is resumed in the cell $\alpha(1-\beta)$; in the fourth period, the cell $(1-\alpha)\beta$ that was unsuccessfully visited in period 2 is revisited, with an expected payoff of zero for that period; and, in the fifth period the remaining $(1-\alpha)(1-\beta)$ cell is visited.

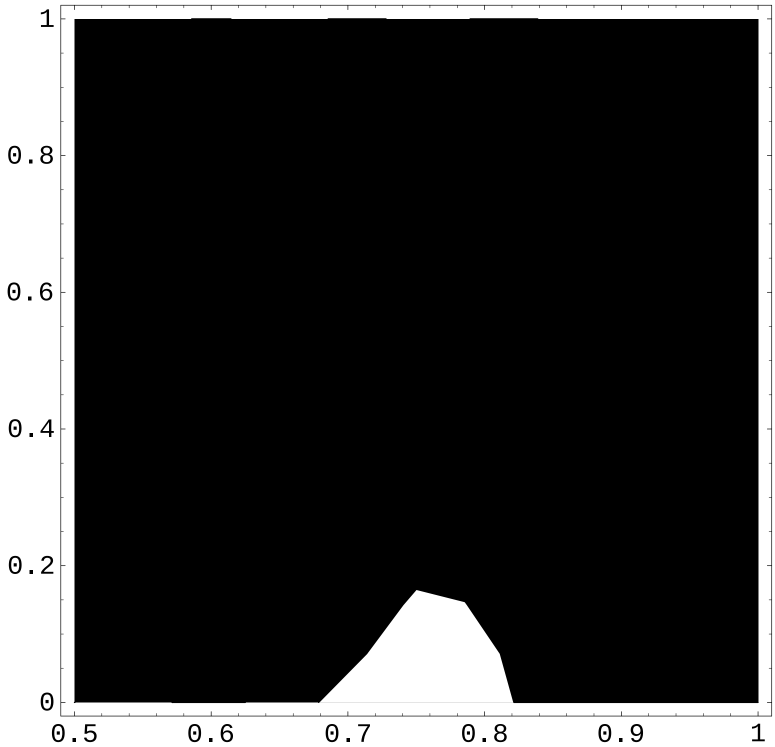
In contrast, the payoff from the exhaustive strategy equals

$$
\begin{aligned}
& 4\int_{\frac{1}{2}}^{1}\int_{\frac{1}{2}}^{1}[\alpha\beta+\delta\alpha(1-\beta)+\delta^2(1-\alpha)\beta+\delta^3(1-\alpha)(1-\beta)]d\alpha d\beta \\
= \quad & \frac{1}{16}[9+3\delta+3\delta^2+\delta^3]
\end{aligned}
$$

We can plot the difference of the two payoffs as a function of c and δ as follows. The plot reveals that for most combinations of c and δ, the exhaustive search strategy is the one with the higher payoff.


0
-0.02
-0.04
-0.06
0.6
0.8
1
0
0.25
0.5
0.75
1

As can be seen in the following contour plot, however, there is a region, indicated in white, in which the payoff from modified exhaustive search exceeds that from exhaustive search. In the black region, the payoff difference between modified exhaustive search and exhaustive search is negative.



5 Results

In this section we generalize the intuition from our $2 \times 2 \times 2$ independent-uniform example to games with an arbitrary finite number of players, an arbitrary finite (not necessarily identical) number of actions per player, an infinite time horizon and a rich class of signal distributions, while maintaining the assumption of signal independence across players.

We begin with the simple observation that players in our game are unable to implement *ex post*-optimal search.

Observation *The full-information optimal strategy (i.e.* ex post-*optimal search) is not a Nash equilibrium strategy.*

To see this, observe that for almost all signal vectors of player i there exists a positive probability set of signal vectors of others players such that the full-information optimal strategy has player i switch in period two. At the same time, there is a positive probability set of signal vectors for which the full-information optimal strategy prescribes that player i does not switch in period two. This behavior cannot be achieved in equilibrium since player i's behavior can only depend on his own information.

Ex post optimal search cannot be implemented because players do not know the signal strengths of other players. In contrast, there is a large class of equilibria in strategies that condition only on the rank order of signals not their value. We refer to such strategies as *ordinal strategies,* and to equilibria in ordinal strategies as *ordinal equilibria.* Some ordinal equilibria are appealing because they induce a novel action profile as long as such profiles are available, i.e. they search exhaustively. We will see, however, that not all ordinal equilibria are exhaustive equilibria and that there is a multitude of Pareto-ranked exhaustive equilibria.

We next formally define ordinal strategies and characterize the set of equilibria in ordinal strategies. Let a_i denote player i's action in the action profile a. Define $r(a_i)$ as the rank of player i's action a_i in descending order of his observed probabilities, i.e. a higher value of $r(a_i)$ indicating a higher probability of a_i being part of a success profile, breaking ties in an arbitrary manner if necessary. Recall the convention that $a_{ij}(\omega_i)$ is player i's jth-ranked action given his signal ω_i and let $h_{it}(\omega_i)$ be player i's private history of such actions until period t. Then a strategy f_i of player i is *ordinal* if $f_i(h_{it}(\omega_i), \omega_i) = f_i(h_{it}(\omega_i), \omega_i')$ for all ω_i, ω_i'. Thus, in an ordinal strategy, the only role of player i's signal is to determine the labeling of his actions according to their rank. A profile f is ordinal if it is composed of ordinal strategies; otherwise, it is *cardinal.* For any ordinal profile f, let $a^t(f)$ denote the profile of actions that the strategy f induces in period t. Similarly, define $A^t(f) := \{a \in A | a^\tau(f) = a \text{ for some } \tau \leq t\}$ as the set of all profiles that the ordinal strategy f induces before period $t+1$.

Notice that if an ordinal strategy profile after some time prescribes an action profile from which a player can reach an unused profile in which his own action has a higher rank, then this player can unilaterally deviate and improve the search order. Conversely, in an ordinal strategy profile in which no player can ever unilaterally reach an unused profile in which his own action has a higher rank than the one prescribe by the profile, no player can deviate to a strategy that would improve the search order. Hence, we have the following characterization of equilibria in ordinal strategies.

Proposition 3 *A profile of ordinal strategies f is a Nash equilibrium if and only if for all players i and all times t there does not exist $a_i \in A_i$ such that (i) $a_i \neq a_i^t(f)$, (ii) $(a_i, a_{-i}^t(f)) \in A \setminus A^{t-1}(f)$, and (iii) $r(a_i) > r(a_i^t(f))$.*

Proof: We begin by proving that any such profile f is an equilibrium. Suppose not. Then there exists a player i, a set of signals S that has positive probability, and a pure strategy f_i' such that $\pi_i(f_i', f_{-i}; \omega_i) > \pi_i(f; \omega_i)$ for all types $\omega_i \in S$. Without loss of generality, we may assume that none of the signals in S require tie breaking in order to determine the probability rank of player i's actions. Consider a type ω_i in S. Let τ be the first period in which $a^\tau(f_i', f_{-i}) \neq a^\tau(f)$. There are two possibilities: Either $a^\tau(f_i', f_{-i}) \in A^{t-1}(f)$, or $a^\tau(f_i', f_{-i}) \in A \setminus A^{t-1}(f)$ and $r(a^\tau(f_i', f_{-i})) < a^\tau(f)$.

Let $\theta > \tau$ be the first period in which $a^\tau_{-i}(f) = a^\theta_{-i}(f)$ whenever such a θ exists. Consider a strategy f_i'' with

$$\begin{aligned} a_i^t(f_i'') &= a_i^t(f_i') \;\; \forall t \neq \tau, \theta \\ a_i^\tau(f_i'') &= a_i^\tau(f_i) \\ a_i^\theta(f_i'') &= a_i^\tau(f_i'). \end{aligned}$$

Evidently, this raises the probability of finding a success in period τ by the same amount that it lowers it in period θ. Because of discounting this raises the payoff of type $\omega_i \in S$.

If there is no θ with $a^\tau_{-i}(f) = a^\theta_{-i}(f)$, replace f_i with a strategy f_i'' such that

$$\begin{aligned} a_i^t(f_i'') &= a_i^t(f_i') \;\; \forall t \neq \tau \\ a_i^\tau(f_i'') &= a_i^\tau(f). \end{aligned}$$

Evidently, also in this case, the payoff of type ω_i is increased.

Iterating this procedure generates a sequence of action profiles that converges to $a^t(f)$. Furthermore the payoff of type $\omega_i \in S$ is non-decreasing at each step of the iteration, contradicting the assumption that f_i' induces a higher payoff for type ω_i than f_i against f_{-i}. This shows that the strategies described in the statement of the proposition indeed form Nash equilibria.

Conversely, suppose the strategy f is an ordinal Nash equilibrium, but there is a player i, an action a_i and a time τ such that conditions (i) – (iii) are satisfied. Let $\theta > \tau$ be the first period in which $a^\tau_{-i}(f) = a^\theta_{-i}(f)$ whenever such a θ exists. Consider a deviation by player i to the strategy f_i' with

$$\begin{aligned} a_i^t(f_i') &= a_i^t(f_i) \;\; \forall t \neq \tau, \theta \\ a_i^\tau(f_i') &= a_i \\ a_i^\theta(f_i') &= a_i^\tau(f_i) \end{aligned}$$

If there is no period $\theta > \tau$ with $a^\tau_{-i}(f) = a^\theta_{-i}(f)$, consider the deviation by player i to the strategy f_i' with

$$\begin{aligned} a_i^t(f_i') &= a_i^t(f_i) \;\; \forall t \neq \tau \\ a_i^\tau(f_i') &= a_i \end{aligned}$$

Evidently, in either case the deviation raises the probability of finding a success in period τ by the same amount that it lowers it in period θ. Because of discounting this raises the payoff of player i for a set of types that has positive probability.

$\square$

	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$
$a_{1,1}$	1	2	3
$a_{1,2}$	4	5	6

FIGURE 1

Observe that the equilibria characterized in Proposition 3 include (i) equilibria in which all profiles are examined without repetition, (ii) equilibria in which search stops before all profiles have been examined, and (iii) infinitely many Pareto-ranked equilibria in which search is temporarily suspended and then resumed. For example, consider the matrix of action profiles in the stage game in FIGURE 1. In the figure $a_{i,j}$ denotes the j-th action of player i, wlog ranked in the order of the corresponding signals, i.e. if we denote by $\alpha_{i,j}$ the probability that the j-th action of player of player i is part of a success profile, then $\alpha_{i,j} \geq \alpha_{i,j+1}$ for all i and j. For convenience, the six action profiles have been numbered. Then, (i) there is an equilibrium in which the profile labelled t is played in period $t = 1, \ldots, 6$, (ii) another equilibrium in which the profile labelled t is played in period $t = 1, \ldots, 4$ after which profile 1 is played forever, and (iii), for any $k > 0$ there is an equilibrium in which the profile labelled t is played in period $t = 1, \ldots, 4$ after which profile 1 is played for k period followed by play of profiles 5 and 6.

We next show that for some class of distributions one can improve on the best ordinal equilibrium. Say that a player's signal distributions has a *mass point at certainty* if there is positive probability that he receives a signal that singles out one of his actions as the one that is part of a success profile. Similarly, say that a player's signal distributions has a *mass point at indifference* if there is positive probability that he receives a signal that assigns equal probability to each of his actions as being part of a success profile. Denote by E_i^C the event that i is certain and by E_i^I the event that he is indifferent.

Lemma 1 *If all players' signal distributions have mass points at certainty and at indifference, the optimal equilibrium is cardinal.*

Proof: In any ordinal equilibrium f, there will be one player, i, who switches in period two, and another player, j, who does not switch in period two. Modify the behavior of these two players as follows: Let i never switch from his first-period action when he is certain. Have j switch in period two to his action a_{j2} (in an ordinal equilibrium, it is without loss of generality to always have him use the action a_{j1} in period 1) when he is indifferent. Have j otherwise not change his behavior, except that in the period $\tau > 2$ with $a^{\tau}(f) = (a_{j2}, a^2_{-j}(f))$, he takes the action a^2_{j1}, instead of a^2_{j2}. Formally, define f' such that $f'_{-ij} = f_{-ij}$, i.e. f coincides with f' for all players other than i and j, and

$$\begin{aligned}
a_i^t(f_i') &= a_i^t(f_i) \ \forall \omega_i \in \Omega_i \setminus E_i^C, t \\
a_i^t(f_i') &= a_i^1(f_i) \ \forall \omega_i \in E_i^C, t \\
a_j^t(f_j') &= a_j^t(f_j) \ \forall \omega_j \in \Omega_j \setminus E_j^I, t \\
a_j^2(f_j') &\neq a_j^1(f_j) \ \forall \omega_j \in E_j^I \\
a_j^{\tau}(f_j') &= a_j^1(f_j) \ \forall \omega_j \in E_j^I \\
a_j^t(f_j') &= a_j^t(f_j) \ \forall \omega_j \in E_j^I, \ t \neq 2, \tau
\end{aligned}$$

There are four possible cases: (1) If player i is uncertain and player j is not indifferent, then the sequence in which cells are examined under the modified strategy profile f' is the same as in the original equilibrium f, and therefore payoffs are the same as well. (2) If player i is certain and player j is not indifferent, then player i is using a dominant action and all other players are following the same behavior as under f_{-i}. Consequently, the expected payoff cannot be lower than from all players using strategy f. (3) If player i is uncertain and player j is indifferent, the only effect of changing from f to f' is that the order in which two cells are visited is reversed. Furthermore, these cells are only distinguished by player j's action and player j is indifferent. Hence payoffs are unchanged in this case. (4) If player i is certain and player j is indifferent, the cell examined in period two has a positive success probability under f', whereas that probability is zero under f. Furthermore, since player i is using a dominant action, and all players other than players i and j do not change their behavior, the overall effect of switching from f to f' is to move the examination of higher probability profiles forward. Therefore, in this case the expected payoff increases. □

Next, we show that the ability to improve on the best ordinal equilibrium does not critically depend on the distribution of signals having mass points.

Proposition 4 *Let F be any distribution of signals that satisfies independence across players and that has a positive density f. Then there exists a sequence of distributions F_n with positive densities f_n and an $N > 0$ such that F_n converges weakly to F and for all $n > N$ the optimal Nash equilibrium is cardinal.*

Proof: Let $e_{ij} \in \Omega_i$ be the signal for player i that assigns probability one to the jth action of player i being part of a success profile, and let $z_i \in \Omega$ be the signal that assigns probability one to the signal that all of player i's actions are equally likely to be part of a success profile. Define E_i to be the distribution of player i's signals that assigns probability one to the set of signals $\{e_{i1}, \ldots, e_{iJ(i)}, z_i\}$ and equal probability to all signals in that set.

Let $\lambda_n \in (0, 1)$ and $\lambda_n \to 0$. Define $G_{n,i} = \lambda_n E_i + (1 - \lambda_n)F_i$, and $G_n = \prod_{i=1}^{I} G_{n,i}$ Then $\{G_n\}_{n=1}^{\infty}$ is a sequence of distributions converging weakly to F, denoted $G_n \xrightarrow{w} F$, where each G_n has mass points at indifference and at certainty. Let $E_{i,k}$ be the distribution of player i's signals that assigns probability one to the set of signals $\tilde{\Omega}_i \subset \Omega_i$ that are within distance $\frac{1}{k}$ from the set $\{e_{i1}, \ldots, e_{iJ(i)}, z_i\}$ and that is uniform on $\tilde{\Omega}_i$. Define $H_{n,i,k} = \lambda_n E_{i,k} + (1 - \lambda_n)F_i$, and let $H_{n,k} = \prod_{i=1}^{I} H_{n,i,k}$. Then $\{H_{n,k}\}_{k=1}^{\infty}$ is a sequence of distribution functions with $H_{n,k} \xrightarrow{w} G_n$, where each $H_{n,k}$ has an everywhere positive density.

Clearly, candidates for an optimal ordinal strategy for $H_{n,k}$ have to be exhaustive. Since there are only finitely many paths of play for exhaustive strategies, an optimal ordinal strategy σ_n^k for $H_{n,k}$ exists. Finiteness of the set of play paths of exhaustive strategies also implies that there is a subsequence of $\{H_{n,k}\}_{k=1}^{\infty}$ for which (after reindexing) the path of play induced by $\{\sigma_n^k\}_{k=1}^{\infty}$ is constant. From now on consider this subsequence, and pick a strategy σ_n that induces this path of play.

Given a signal realization ω, denote player i's expected payoff from the strategy profile σ by $v_i(\sigma, \omega)$. Then, for any strategy profile σ and signal distribution Φ, player i's expected payoff $U_i(\sigma, \Phi)$ is

$$U_i(\sigma, \Phi) = \int v_i(\sigma, \omega) d\Phi(\omega).$$

Let $1_{\{\sigma,a,t\}}$ be the indicator of the event that profile a is visited for the first time in period t under strategy σ and $P(a|\omega)$ the probability that the profile a is a success given the signal vector ω. Then, for an ordinal strategy $\tilde{\sigma}$, player i's payoff for a fixed ω has the form

$$v_i(\tilde{\sigma}, \omega) = \sum_{t=1}^{\infty} \delta^{t-1} 1_{\{\tilde{\sigma},a,t\}} P(a|\omega).$$

Here $P(a|\omega)$ is a polynomial in ω and therefore varies continuously with ω. Since $\tilde{\sigma}$ is ordinal, the indicator function does not vary with ω (recall that is ω_i passes through a point of indifference between two action, they simply become relabeled). Taken together, these observations imply that $v_i(\tilde{\sigma}, \omega)$ is continuous in ω. Hence, by weak convergence of $H_{n,k}$ to G_n, we have

$$U_i(\tilde{\sigma}, H_{n,k}) \to U_i(\tilde{\sigma}, G_n),$$

for any ordinal strategy $\tilde{\sigma}$. Therefore, σ_n must be an optimal ordinal strategy for G_n.

Denote by σ'_n the improvement strategy for σ_n, given G_n as constructed in the proof of Lemma (1). For any given σ'_n, σ_n and $\epsilon > 0$, we can define the strategy σ^ϵ_n as follows:

$$\sigma^\epsilon_{i,n}(\omega_i) = \begin{cases} \sigma_{i,n}(\omega_i) \text{ if } |\omega_i - e_{ij}| > \epsilon \text{ and } |\omega_i - z_i| > \epsilon \\ \frac{\epsilon - x}{\epsilon}\sigma'_{i,n}(z_i) + \frac{x}{\epsilon}\sigma_{i,n}(\omega_i) \text{ if } |\omega_i - z_i| = x \leq \epsilon \\ \frac{\epsilon - x}{\epsilon}\sigma'_{i,n}(e_{ij}) + \frac{x}{\epsilon}\sigma_{i,n}(\omega_i) \text{ if } |\omega_i - e_{ij}| = x \leq \epsilon \end{cases}$$

Note that the payoff $v_i(\sigma^\epsilon_n, \omega)$ is a continuous function of the signal vector ω. Hence, weak convergence implies that

$$U_i(\sigma^\epsilon_n, H_{n,k}) \to U_i(\sigma^\epsilon_n, G_n) \text{ as } k \to \infty.$$

By construction, we have

$$U_i(\sigma'_n, G_n) > U_i(\sigma_n, G_n),$$

and

$$U_i(\sigma^\epsilon_n, G_n) \to U_i(\sigma'_n, G_n) \text{ as } \epsilon \to 0.$$

Note that the payoff $v_i(\sigma_n, \omega)$ is a continuous function of the signal vector ω. Hence, weak convergence implies that

$$U_i(\sigma_n, H_{n,k}) \to U_i(\sigma_n, G_n) \text{ as } k \to \infty.$$

Combining these observations, we conclude that for any n, we can find $k(n)$ and $\tilde{\epsilon}$ such that

$$U_i(\sigma^{\tilde{\epsilon}}_n, H_{n,k(n)}) > U_i(\sigma_n, H_{n,k(n)}).$$

To conclude, simply let $F_n = H_{n,k(n)}$. □

References

Alpern, S. [1976], "Hide and Seek Games," Seminar, Institut für höhere Studien, Wien, 26 July.

Alpern, S. [2002], "Rendezvous Search: A Personal Prespective," *Operations Research,* **50**, .

Alpern, A. and S. Gal [2003], *The Theory of Search Games and Rendezvous,* Dordrecht, The Netherlands: Kluwer Academic Publishers.

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

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. and A.M. Franco [2005], "Decentralized Learning from Failure," forthcoming in *Journal of Economic Theory.*

Blume, A., J. Duffy and A.M. Franco [2004], "Decentralized Organizational Learning: An Experimental Investigation," University of Iowa and University of Pittsburgh Working Paper.

Chamley, C.P. [2004], *Rational Herds: Economic Models of Social Learing* Cambridge, UK: Cambridge University Press.

Crawford, V. and H. Haller [1990], "Learning how to Cooperate: Optimal Play in Repeated Coordination Games," *Econometrica,* **58**, 571-595.

Gossner, O. and N. Vieille [2003], "Strategic Learning in Games with Symmetric Information," *Games and Economic Behavior,* **42**, 25-47.

McLennan, A. [1998], "Consequences of of the Condorcet Jury Theorem for Beneficial Information Aggregation by Rational Agents," *American Political Science Review* , **92**, 413-418.

Wiseman, T. [2005], "A Partial Folk Theorem for Games with Unknown Payoff Distributions," *Econometrica,* **73**, 629-645.