Conjectural Equilibrium in Multiagent Learning

MICHAEL P. WELLMAN   wellman@umich.edu
JUNLING HU   junling@umich.edu
University of Michigan, Ann Arbor, MI 48109-2110

Editors: Michael Huhns and Gerhard Weiß

Abstract. Learning in a multiagent environment is complicated by the fact that as other agents learn, the environment effectively changes. Moreover, other agents' actions are often not directly observable, and the actions taken by the learning agent can strongly bias which range of behaviors are encountered. We define the concept of a *conjectural equilibrium*, where all agents' expectations are realized, and each agent responds optimally to its expectations. We present a generic multiagent exchange situation, in which competitive behavior constitutes a conjectural equilibrium. We then introduce an agent that executes a more sophisticated strategic learning strategy, building a model of the response of other agents. We find that the system reliably converges to a conjectural equilibrium, but that the final result achieved is highly sensitive to initial belief. In essence, the strategic learner's actions tend to fulfill its expectations. Depending on the starting point, the agent may be better or worse off than had it not attempted to learn a model of the other agents at all.

Keywords: multiagent learning, conjectural equilibrium, market game

This article includes some material previously reported at
ICMAS-96 (Hu & Wellman, 1996).

1. Introduction

Machine learning researchers have recently begun to investigate the special issues that multiagent environments present to the learning task. Contributions in this journal issue, along with recent workshops on the topic (Grefenstette et al, 1996, Sen, 1996, Sen, 1997), have helped to frame research problems for the field. Multiagent environments are distinguished in particular by the fact that as the agents learn, they change their behavior, thus effectively changing the environment for all of the other agents. When agents are acting and learning simultaneously, their decisions affect (and limit) what they subsequently learn.

1.1. Learning and Equilibrium

The changing environment and limited ability to learn the full range of others' behavior presents pitfalls, both for the individual learning agent and for the designer of multiagent learning methods. For the latter, it is not immediately obvious even how to define the goal of the enterprise. Is it to optimize the effectiveness of an individual learning agent across a range of multiagent configurations, or to optimize the joint effectiveness of a configuration of learning agents? Of course, either problem may predominate depending on the circumstance. In any case, we require a framework for characterizing a multiagent learning process, and analyzing the behaviors of alternative learning regimes.

We argue that a central element of such a multiagent learning framework is an *equilibrium concept*, that is, a characterization of some steady-state balance relationship among the agents. This follows by direct analogy from the static knowledge (i.e., no learning) case. In single-agent decision theory, the agent's problem is to maximize its utility. This remains true in the multiagent (i.e., game-theoretic) case, but there all the agents are simultaneously optimizing. The equilibrium (consistent joint optimization) thus represents the logical multiagent extension of individual optimization. Although from any individual agent's perspective the other agents may well be treated as part of the environment, a decision on the analyst's part to accord all of them *agent* status (i.e., to treat the system as multiagent) imposes an essential symmetry on the problem.

Note that equilibrium is an idealization of multiagent behavior, just as optimization is an idealization of single-agent behavior. Whether or not we actually expect a complicated system to reach equilibrium (or analogously, an individual to optimize successfully), it is quite useful for analysts to understand what these equilibria are. Any nonequilibrium gives at least one agent a motivation to change, just as a nonoptimum is a cause for change in the single-agent case.

Game theory bases its solutions on equilibrium actions (or more generally, *policies*). An agent behaving within an equilibrium is often explained in terms of the agent's *beliefs* about the types or policies of other agents. How agents reach such beliefs through repeated interactions is what game theorists mean by *learning* (Milgrom & Roberts, 1991), and that is the sense of the term we adopt as well.

The distinction between learning and nonlearning agents, for our purposes, is simply that the former change their beliefs, whereas the latter's beliefs are static.[1] Thus, a learning regime defines a dynamic process, and the outcomes achieved in likely trajectories of such processes distinguish the effectiveness of alternative regimes. In the multiagent context, we are interested particularly in whether a learning regime leads to equilibrium behavior, and if so, then how, and when, and which one.

In the approach to multiagent learning proposed here, we characterize an agent's belief process in terms of *conjectures* about the effects of their actions. We define learning in terms of the dynamics of conjectures, and equilibrium in terms of consistency of conjectures within and across agents.

1.2. *A Study in Conjectural Equilibrium*

We proceed in the next section to define our basic solution concept, that of *conjectural equilibrium*. In the sequel, we investigate the concept by exploring a simple multiagent environment representing a generic class of exchange interactions. We identify some interesting phenomena in this context that—while specific to the particulars of the environment and agent assumptions—we suspect to be prevalent in many other circumstances. Following the empirical analysis of this particular environment, we undertake a theoretical analysis that establishes some equilibrium and convergence properties within a somewhat more general setting.

In our basic setup, one class of agents (called *strategic*) attempt to learn models of the others' behavior, while the rest learn a simple reactive policy. We find the following:

1. The system reliably converges to a conjectural equilibrium, where the strategic agents' models of the others are fulfilled, all the rest correctly anticipate the resulting state, and each agent behaves optimally given its expectation.

2. Depending on its initial belief, a strategic agent may be better or worse off than had it simply behaved reactively like the others.

The apparent paradox in this situation is that the learning itself is highly effective: the other agents behave exactly as predicted given what the agent itself does. The paradox is easily resolved by noting that the learned model does *not* correctly predict what the result would be if the agent selected an alternative action. Nevertheless, it is perhaps surprising how easy it is for the agent to get trapped in a suboptimal equilibrium, and that the result is often substantially worse than if it had not attempted to learn a model at all.

We refer to the above situation as *self-fulfilling bias*, because the revisions of belief and action by the agent reinforce each other so that an equilibrium is reached. Here bias is defined as in the standard machine learning literature—the preference for one hypothesis over another, beyond mere consistency with the examples (Russell & Norvig, 1995). In reinforcement learning, the initial hypothesis is a source of bias, as is the hypothesis space (in multiagent environments, expressible models of the other agents). The combination of a limited modeling language (in our experiments, linear demand functions) with an arbitrarily assigned initial hypothesis strongly influences the equilibrium state reached by the multiagent system.

Much early work on multiagent learning has investigated some form of reinforcement learning (e.g., (Tan, 1993, Weiß, 1993)). The basic idea of reinforcement learning is to revise beliefs and policies based on the success or failure of observed performance (Kaelbling, Littman & Moore, 1996). The complication in a multiagent environment is that the rewards to alternative policies may change as other agents' beliefs evolve simultaneously (Claus & Boutilier, 1998, Ono & Fukumoto, 1996).

2. Conjectural Equilibrium

In game-theoretic analysis, conclusions about equilibria reached are based on assumptions about what knowledge the agents have. For example, choice of iterated undominated strategies follows from common knowledge of rationality and the game setup (Brandenburger, 1992). In the standard game-theoretic model of complete information (Fudenberg & Tirole, 1991, Gibbons, 1992), the joint payoff matrix is known to every agent. Uncertainty can be accommodated in the game-theoretic concept of *incomplete information*, where agents have probabilities over the payoffs of other agents.

A learning model is an account of how agents form such beliefs. Notice that the beliefs need not be expressed in terms of other agents' options and preferences. In particular, ignorance about other agents might be captured more directly, albeit abstractly, as uncertainty in the effects of the agent's own actions.[2]

Consider an n-player one-stage game $G = (A, U, S, s)$. $A = A^1 \times \cdots \times A^n$ is the joint action space, where A^i is the action space for agent i. $U = (U^1, \ldots, U^n)$ represents the agent utility functions. $S = S^1 \times \cdots \times S^n$ is the state space, where S^i is the part of the state relevant to agent i. A utility function U^i is a map from the agent's state space to real

numbers, $U^i : S^i \to \Re$, ordering states by preference. We divide the state determination function $s : A \to S$, into components, $s^i : A \to S^i$, allowing each agent's part of the state to depend on the entire joint action. Each agent knows only its own utility function, and the actions chosen by each agent are not directly observable to the others.

Each agent has some belief about the state that would result from performing its available actions. We represent this by a function, $\tilde{s} : A^i \to S^i$, where $\tilde{s}^i(a^i)$ represents the state that agent i believes would result if it selected action a^i. Agent i chooses the action $a^i \in A^i$ it believes will maximize its utility.[3]

We are now ready to define our equilibrium concept.

Definition 1.   In game G defined above, a configuration of beliefs $(\tilde{s}^{1*}, \ldots, \tilde{s}^{n*})$, together with a joint action $a^* = (a^{1*}, \ldots, a^{n*})$, constitutes a *conjectural equilibrium* if, for each agent i,

$$\tilde{s}^{i*}(a^{i*}) = s^{i*}(a^{1*}, \ldots, a^{n*}),$$

where $a^{i*} \in A^i$ maximizes $U^i(\tilde{s}^{i*}(a^i))$.

If the game is repeated over time, the agents can learn from prior observations. Let $a^i(t)$ denote the action chosen by agent i at time t. The state at time t, $\sigma(t)$, is determined by the joint action,

$$\sigma(t) = s(a^1(t), \ldots, a^n(t)).$$

We could incorporate environmental dynamics into the model by defining state *transitions* as a function of joint actions plus the current state. We refrain from taking this step in order to isolate the task of learning about other agents from the (essentially single-agent) problem of learning about the environment.[4] In consequence, our framework defines a repeated game where agents are myopic, optimizing only with respect to the next iteration.

The dynamics of the system are wholly relegated to the evolution of agents' conjectures. At the time agent i selects its action $a^i(t)$, it has observed the sequence $\sigma(0), \sigma(1), \ldots, \sigma(t-1)$. Its beliefs, $\tilde{s}^i$, therefore, may be conditioned on those observations (as well as its own prior actions), and so we distinguish beliefs at time t with a subscript, $\tilde{s}^i_t$. We say that a learning regime *converges* if $\lim_{t \to \infty}(\tilde{s}^1_t, \ldots, \tilde{s}^n_t)$ is a conjectural equilibrium. Our investigation below shows that some simple learning methods are convergent in a version of the game framework considered above.

A Nash equilibrium for game G is a profile of actions $(a^1, \ldots, a^n)$ such that for all i, a^i maximizes $U^i(s^i(a^i, a^{-i}))$. Our notion of conjectural equilibrium is substantially weaker, as it allows the agent to be wrong about the results of performing alternative actions. Nash equilibria are trivially conjectural equilibria where the conjectures are consistent with the equilibrium play of other agents. As we see below, competitive, or Walrasian, equilibria are also conjectural equilibria.

The concept of *self-confirming equilibrium* (Fudenberg & Levine, 1993) is another relaxation of Nash equilibrium which applies to a situation where no agent ever observes actions of other agents contradicting its beliefs. Conjectures are on the play of other agents, and must be correct for all reachable information sets. This is stronger than conjectural

equilibrium in two respects. First, it applies at each stage of an extensive form game, rather than for single-stage games or in the limit of a repeated game. Second, it takes individual actions of other agents as observable, whereas in our framework the agents observe only resulting state.

The basic concept of conjectural equilibrium was first introduced by Hahn, in the context of a market model (Hahn, 1977). Though we also focus on market interactions, our central definition applies the concept to the more general case. Hahn also included a specific model for conjecture formation in the equilibrium concept, whereas we relegate this process to the learning regime of participating agents.

3. Multiagent Market Framework

We study the phenomenon of self-fulfilling bias in the context of a simple market model of agent interactions. The market context is generic enough to capture a wide range of interesting multiagent systems, yet affords analytically simple characterizations of conjectures and dynamics. Our model is based on the framework of general equilibrium theory from economics, and our implementation uses the WALRAS market-oriented programming system (Wellman, 1993), which is also based on general equilibrium theory.

3.1. General Equilibrium Model

Definition 2. A *pure exchange economy* over m goods, $E \equiv \{\langle X^i, U^i, e^i \rangle \mid i = 1, \ldots, n\}$, consists of n consumer agents, each defined by:

- a *consumption set*, $X^i \subseteq \Re_+^m$, representing the bundles of the m goods that are feasible for i,

- a *utility function*, $U^i : X^i \to \Re$, ordering feasible consumption bundles by preference, and

- an *endowment*, $e^i \in \Re_+^m$, specifying i's initial allocation of the m goods.

For example, each of a collection of software agents may have some endowment of various computational resources, such as processing, storage, and network bandwidth. The amounts of these resources controlled by the agent determine which tasks it can accomplish, and at what performance level. The consumption set would describe the minimal amount of these resources required to remain active, and the utility function would describe the value to the agent of results producible with various amounts of the respective resources.

In an exchange system, agents may improve their initial situations by swapping resources with their counterparts. For instance, one network-bound agent might trade some of its storage for bandwidth, while another might use additional storage obtained to improve the result achievable with even a somewhat reduced amount of processing.[5]

The relative prices of goods govern their exchange. The *price vector*, $P \in \Re_+^m$, specifies a price for each good, observable by every consumer agent. A *competitive* consumer takes the price vector as given, and solves the following optimization problem,

$$\max_{x^i \in X^i} U^i(x^i) \text{ s.t. } P \cdot x^i \le P \cdot e^i. \tag{1}$$

That is, each agent chooses a consumption bundle x^i to maximize its utility, subject to the *budget constraint* that the cost of its consumption cannot exceed the value of its endowment.

A *competitive*—also called *Walrasian*—*equilibrium* is a price vector and associated allocation, $(P^*, (x^1, \dots, x^n))$, such that

1. at price vector P^*, x^i solves problem (1) for each agent i, and

2. the markets clear: $\sum_{i=1}^n x^i = \sum_{i=1}^n e^i$.

It is sometimes more convenient to characterize the agents' actions in terms of *excess demand*, the difference between consumption and endowment,

$$z^i = x^i - e^i,$$

and to write the market clearing condition as $\sum_{i=1}^n z^i = 0$. The *excess demand set* for consumer i is $Z^i = \{z^i \in \Re^m \mid e^i + z^i \in X^i\}$.

A basic result of general equilibrium theory (Takayama, 1985) states that if the utility function of every agent is quasiconcave and twice differentiable, then E has a unique competitive equilibrium.[6]

Observe that any competitive equilibrium can be viewed as a conjectural equilibrium, for an appropriate interpretation of conjectures. The action space A^i of agent i is its excess demand set, Z^i. Let the state determination function s return the desired consumptions if they satisfy the respective budget constraints with respect to the market prices, and zero otherwise. Utility function U^i simply evaluates i's part of the allocation. The agents' conjectures amount to accurately predicting the budget constraint, or equivalently, the prices. In competitive equilibrium, each agent is maximizing with respect to its perceived budget constraint, and the resulting allocation is as expected. Thus, the conditions for conjectural equilibrium are also satisfied.

3.2. *Iterative Bidding Processes*

The basic definition of competitive behavior (1) implicitly assumes that agents are *given* the prices used to solve their optimization problem. But it is perhaps more realistic for them to form their own expectations about prices, given their observations and other knowledge they may have about the system. Indeed, the dynamics of an exchange economy can be described by adding a temporal component to the original optimization problem, rewriting (1) as

$$\max_{x^i(t)} U^i(x^i(t)) \text{ s.t. } \tilde{P}^i(t) \cdot x^i(t) \le \tilde{P}^i(t) \cdot e^i(t), \tag{2}$$

where $x^i(t)$ denotes i's demand at time t, and $\tilde{P}^i(t)$ denotes its *conjectured* price vector at that time.[7]

A variety of methods have been developed for deriving competitive equilibria through repeated agent interactions. In many of these methods, the agents do not interact directly, but rather indirectly through auctions. Agents submit bids, observe the consequent prices, and adjust their expectations accordingly.
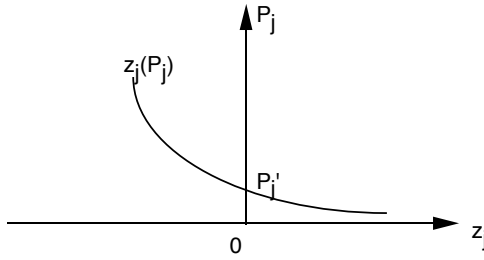
Figure 1. An aggregate excess demand curve for good j. P'_j is the market clearing price.

Different ways of forming the expected price $\tilde{P}(t)$ characterize different varieties of agents, and can be considered alternative learning regimes. For example, the *simple competitive agent* takes the latest actual price as its expectation,

$$\tilde{P}^i(t) = P(t-1). \tag{3}$$

More sophisticated approaches are of course possible, and we consider one in detail in the next section.

In the classic method of *tatonnement*, for example, auctions announce the respective prices, and agents act as simple competitors. Depending on whether there is an excess or surfeit of demand, the auction raises or lowers the corresponding price. If the aggregate demand obeys *gross substitutability* (an increase in the price of one good raises demand for others, which hence serve as substitutes), then this method is guaranteed to converge to a competitive equilibrium (under the conditions under which it is guaranteed to exist) (Negishi, 1962).

The WALRAS algorithm (Cheng & Wellman, 1998) is a variant of tatonnement. In WALRAS, agent i submits to the auction for good j at time t its solution to (2), expressed as a function of P_j, assuming that the prices of goods other than j take their expected values. In other words, it calculates a *demand function*,

$$x^i_j(\tilde{P}^i_1(t), \ldots, P_j, \ldots, \tilde{P}^i_m(t)).$$

The bid it then submits to the auctioneer is its excess demand for good j,

$$z^i_j(P_j) = x^i_j(\tilde{P}^i_1(t), \ldots, P_j, \ldots, \tilde{P}^i_m(t)) - e^i_j(t).$$

The auctioneer sums up all the agents' excess demands to get an *aggregate excess demand* function,

$$z_j(P_j) = \sum_{i=1}^{n} z^i_j(P_j).$$

Figure 1 depicts an aggregate demand curve. We assume that $z_j(P_j)$ is downward sloping, the general case for normal goods. Given such a curve, the auctioneer determines the price P'_j such that $z_j(P'_j) = 0$, and reports this *clearing price* to the interested agents.

Given the bidding behavior described, with expectations formed as by the simple competitive agent, the WALRAS algorithm is guaranteed to converge to competitive equilibrium, under the standard conditions (Cheng & Wellman, 1998). Such an equilibrium also represents a conjectural equilibrium, according to the definition above. Thus, the simple competitive learning regime is convergent, with respect to both the tatonnement and WALRAS price adjustment protocols.

4.   Learning Agents

As defined above, agents *learn* when they modify their conjectures based on observations. We distinguish alternative learning regimes by the form of the conjectures produced, and the policies for revising these conjectures.

4.1.   Competitive Learning Agents

An agent is *competitive* if it takes prices as given, ignoring its own effect on the clearing process. Formally, in our learning framework, this means that the conjectured prices $\tilde{P}$ do not depend on the agents' own actions—the excess demands they submit as bids. For example, the simple competitive agent described above simply conjectures that the last observed price is correct. This revision policy is given by (3).

Adaptive competitive agents adjust their expectations according to the difference between their previous expectations and the actual observed price,

$$\tilde{P}^i(t) = \tilde{P}^i(t-1) + \gamma \left(P(t-1) - \tilde{P}^i(t-1) \right).$$

This updating method is a kind of reinforcement learning method. The learning parameter, γ, dictates the rate at which the agent modifies its expectations. When $\gamma = 1$, this policy is identical to the simple competitive agent's. Variations on this adaptation, for example by tracking longer history sequences, also make for reasonable conjecture revision policies.

4.2.   Strategic Learning Agents

In designing a more sophisticated learning agent, we must take into account what information is available to the agent. In our market model, the agents cannot observe preference, endowment, or the complete demand functions of other agents. What the agent does observe is the price vector. It also knows the basic structure of the system—the bidding process and the generic properties we assume about demand.

This fragmentary information is not sufficient to reconstruct the private information of other agents. In fact, it provides no individual information about other agents at all. The best an agent can do is learn about the aggregate action it faces.

Because they know how the auctions work, the agents realize that their individual demands can affect the market price. This effect will be significant unless the agent is of negligible

size with respect to the aggregate system. An agent that takes its own action into account in forming its expectation about prices is called *strategic*. For a strategic agent i, $\tilde{P}^i$ is a function of excess demand, $z^i(t)$, and thus i's optimization problem is subject to a nonlinear budget constraint.

In our experiments with strategic learning, we adopt a simple model of an agent's influence on prices. Specifically, the agent assumes that its effect on price is linear for each good j,

$$\tilde{P}^i_j(t) = \alpha^i_j(t) + \beta^i_j(t) z^i_j(t), \text{ where } \beta^i_j(t) \geq 0. \tag{4}$$

As in our usual reinforcement-learning approach, the coefficients are adjusted according to the difference between the expected price and actual price,

$$\alpha^i_j(t+1) = \alpha^i_j(t) + \gamma_1 \left(P_j(t) - \tilde{P}^i_j(t) \right), \tag{5}$$

$$\beta^i_j(t+1) = \beta^i_j(t) + \frac{\gamma_2}{z^i_j(t)} \left(P_j(t) - \tilde{P}^i_j(t) \right), \tag{6}$$

where γ_1 and γ_2 are positive constants.

Thus, by substituting (4) into (2) and omitting the time argument, we obtain the optimization problem of the strategic agent,

$$\max_{z^i} U^i(z^i + e^i) \text{ s.t. } (\alpha^i + \beta^i z^i) \cdot z^i \leq 0. \tag{7}$$

In the appendix, we prove that this problem indeed has a unique solution.

5. Experimental Results

We have run several experiments in WALRAS, implementing exchange economies with various forms of learning agents. Our baseline setup explores the behavior of a single strategic learning agent (as described above), included in a market where the other agents are simple competitors. Additional trials consider different numbers of strategic agents, and varying initial conditions.

Agents in our experiments have logarithmic utility functions,

$$U(x_1, \ldots, x_m) = \sum_j a_j \ln x_j.$$

This utility function is strategically equivalent to the Cobb-Douglas form, which is a standard parametric family often employed for analytic convenience.[8] For the experiments, we set $a_j = 1$ for all j, for all agents.

Because its price conjecture is a function of its action, the strategic agent faces a nonlinear budget constraint, and thus a more complex optimization problem (7). This special form facilitates derivation of first-order conditions, which we solve numerically in our experimental runs to calculate the strategic agent's behavior.

In our simulations, the competitive agents form conjectures by Equation (3). The strategic agent forms conjectures by (4), and revises them given observations according to (5) and (6), with $\gamma_1 = \gamma_2 = \frac{1}{2}$. Agents bid according to the solutions of their optimization problems.

The auctioneer in each market receives bids from agents, and then posts the price that clears its market. The process terminates when the price change from one iteration to the next falls below some threshold.

We performed a series of experiments for a particular configuration with three goods and six agents. The agents' endowments e^i were randomly generated from a uniform distribution, with results displayed in Table 1. Figure 2 presents results for the case where agent 1 behaves strategically, and the rest competitively. Each point on the graph represents one run of this economy, with various settings of the strategic agent's initial conjecture. The vertical axis represents the utility achieved by the strategic agent when the system reaches equilibrium. The horizontal axis represents the strategic agent's starting value for its β coefficient. For comparison, we also ran this configuration with the designated agent behaving competitively, that is, forming expectations independent of its own behavior according to (3). The utility thus achieved is represented by the horizontal line in the graph.

Table 1. Initial endowments for agents in the example experiment.

Agents	Good 1	Good 2	Good 3
Agent 1	231	543	23
Agent 2	333	241	422
Agent 3	43	21	11
Agent 4	33	24	42
Agent 5	431	211	111
Agent 6	12	23	87

As Figure 2 demonstrates, the learning agent can achieve higher or lower payoff by attempting to behave strategically rather than competitively. For $\beta^1(0) < 0.03$, the agent improves utility by learning the strategic model. Greater than that value, the agent would be better off behaving competitively. (We also ran experiments for higher values of $\beta^1(0)$ than shown, and the trend continues. In some other instances of the market game, the strategic agent also does worse than competitive for excessively low values of $\beta(0)$.) Intuitively, the initial estimate of the agent's effect on prices moves it toward a demand policy that would fulfill this expectation.

The utility achieved by the other agents also depends on the initial β of the strategic agent. Figure 3 depicts the results for the competitive agents, using as a measure the ratio of utility achieved when agent 1 is strategic to that achieved when it is competitive. For these agents, we find that two (3 and 5) are better off when agent 1 behaves strategically, and the rest are worse off. Moreover, their resulting utilities are monotone in $\beta^1(0)$. Note that the agents that do better have endowment profiles (see Table 1) relatively similar to agent 1, and thus agent 1's effect on the price turns out to their benefit. The other agents have relatively differing endowment profiles, and thus opposing interests.

In general, results need not be so uniform. We have observed cases where competitive agents do not perform uniformly better or worse as another becomes strategic, and indeed
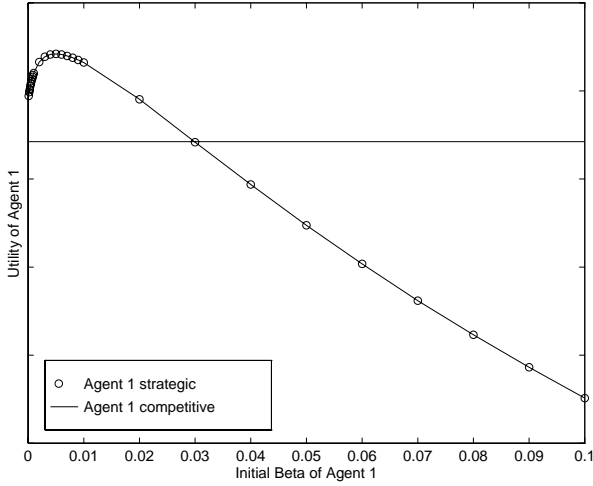
Figure 2. Utility achieved by the strategic agent, as a function of $\beta^1(0)$. (Since utility is only ordinally scaled, the shape of the curve and degrees of utility difference are not meaningful. Hence, we do not report numeric values on the vertical axis.)

it is possible that aggressive strategic behavior can even make all agents worse off. In contrast, it is not possible that strategic behavior can simultaneously make all better off, as competitive equilibria are guaranteed to be Pareto efficient.
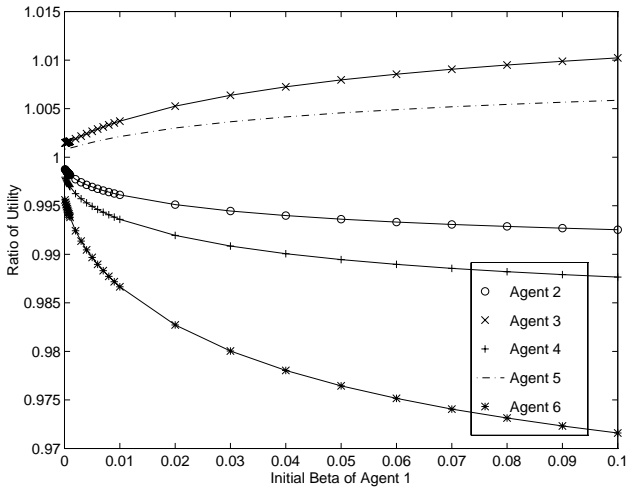

Figure 3. Performance of the competitive agents, as a function of $\beta^1(0)$. The vertical axis measures the ratio of utility when agent 1 is strategic versus when it is competitive.

As we increase the number of competitive agents, the general patterns of Figures 2 and 3 still hold. We also ran experiments with multiple strategic agents in the system. For example, Figure 4 compares strategic agent 1's performance profile for the cases where agent 3 behaves strategically and competitively. In most of our experiments, the system reliably converges to a conjectural equilibrium, although the particular equilibrium reached depends on the initial model of the strategic learning agents.[9] The exceptions are cases where the combined power of the strategic agents is relatively large, opening the possibility that markets will not clear for significantly erroneous conjectures. This situation is explained in more detail in Section 6.2.
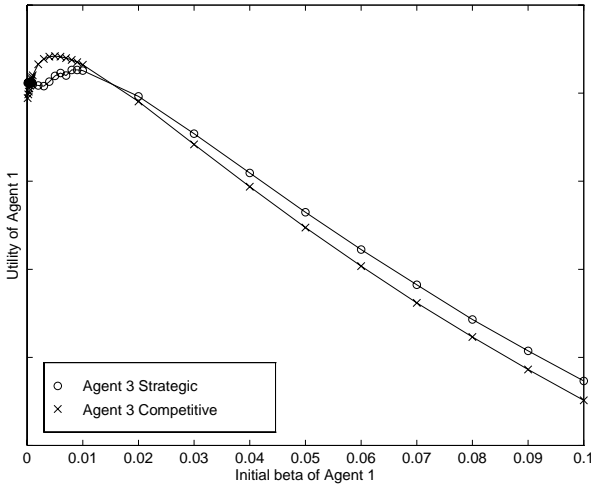

Figure 4. Utility achieved by strategic agent 1, as a function of $\beta^1(0)$, with agent 3 strategic and competitive, respectively.

6. Theoretical Analysis

The sensitivity of outcomes to initial conjectures arises from lack of information. When an agent has incomplete knowledge about the preference space of other agents, its interaction with them may not reveal their true preferences even over time. Nevertheless, agents adopting myopic decision rules (e.g., best response) may well achieve conjectural equilibrium anyway.

In this section, we specialize the concept of conjectural equilibrium to the multiagent exchange setting. We define the *market conjectural equilibrium*, and discuss its existence and multiplicity for particular classes of learning agents. We then consider the dynamics of strategic learning in this framework, and conditions for convergence to conjectural equilibrium.

6.1. Market Conjectural Equilibrium

Our experimental analysis considered agents whose conjectures were either constant (competitive) or linear (strategic) functions their actions. Using Hahn's notion of a *conjecture function* (Hahn, 1977), we provide some more general notation for characterizing the form of an agent's conjectures.

Definition 3. The *conjecture function*, $C^i : \Re^m \longrightarrow \Re^m_+$, specifies the price system, $C^i(z^i)$, conjectured by consumer i to result if it submits excess demand z^i.

Note that C^i defines a conjecture about prices, whereas conjectural equilibrium is defined in terms of agent's conjectures about the effects of their actions. In the multiagent exchange setting, actions are excess demands, and an agent's conjecture about the resulting state, $\tilde{s}^i$, is that it will receive its demanded bundle if and only if it satisfies its budget constraint.

$$\tilde{s}^i(z^i) = \begin{cases} z^i & \text{if } C^i(z^i) \cdot z^i \leq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

The *actual* resulting state is as demanded if the aggregate demands are feasible.[10] For all i,

$$s^i(z^1, \ldots, z^n) = \begin{cases} z^i & \text{if } \sum_k z^k \leq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

In conjectural equilibrium, the expected and actual consequences of optimizing behavior coincide.

Definition 4. A *market conjectural equilibrium* for an exchange economy is a point $(C^1, \ldots, C^n)$ such that for all i, $\tilde{s}^i(z^i) = z^i$, where

$$z^i = \arg\max U^i(z^i + e^i) \text{ s.t. } C^i(z^i) \cdot z^i = 0,$$

and $\sum_i z^i \leq 0$.

Intuitively, $C^i(z^i) = P$, where P is the price vector determined by the market mechanism. However, nothing in the definition actually requires that all agents conjecture the same price, as the price is not part of an agent's action or the resulting state (9). It is nevertheless worth noting that equivalent price conjectures with overall feasibility is a sufficient condition for market conjectural equilibrium.

THEOREM 1 *Let E be an exchange economy where all agents are allowed to form arbitrary price conjectures. Then any feasible allocation in which each agent prefers the result to its endowment can be supported by a market conjectural equilibrium in E.*

Proof: Let $z^{*1}, \ldots, z^{*n}$ represent a set of excess demands satisfying the conditions, that is, $z^{*i} \in Z^i$ and $U^i(z^{*i} + e^i) \geq U^i(e^i)$ for all i, and $\sum_i z^{*i} \leq 0$. Consider a z^i that agent i prefers to z^{*i}, that is, $U^i(z^i + e^i) > U^i(z^{*i} + e^i)$. It is easy to construct a conjecture

function for agent i such that $C^i(z^i) \cdot z^i > 0$ for any such z^i, in which case i believes that choosing z^i would violate its budget constraint and therefore result in consumption of e^i. Since $U^i(z^{*i} + e^i) \geq U^i(e^i)$, z^{*i} maximizes utility with respect to the conjecture. ∎

With restrictions on the form of individual conjectures, the set of equilibria may be somewhat constrained, but not very much. More realistic situations account for the fact that agents' conjectures are connected to each others via *prices*.

If prices are observed by the agents in an exchange economy, then conjectures inconsistent with the observed prices represent implausible agent behavior. We can capture the notion of consistency among price conjectures in a stronger equilibrium concept.

Definition 5.   A market conjectural equilibrium $(C^1, \ldots, C^n)$ is *price-ratified* if there exists a price vector P such that at the equilibrium actions,

$$z^i = \arg\max U^i(z^i + e^i) \text{ s.t. } C^i(z^i) \cdot z^i = 0,$$

$C^i(z^i) = P$ for all i.

Because prices are known by agents in typical market settings (albeit often with some delay), price-ratified equilibrium is usually the more relevant concept. Indeed, the equilibria reached in our experiments of Section 5 are all price-ratified. We can now characterize the existence of price-ratified market conjectural equilibria in terms of the allowable conjecture functions.

THEOREM 2 *Suppose E has a competitive equilibrium, and all agents are allowed to form constant conjectures. Then E has a price-ratified market conjectural equilibrium.*

Proof: Let P^* be a competitive equilibrium for E. Then $C^i(z^i) = P^*$, for all $z^i \in Z^i$, $i = 1, \ldots, n$, is a market conjectural equilibrium, ratified by P^*. ∎

THEOREM 3 *Let E be an exchange economy, with all utility functions quasiconcave and twice differentiable. Suppose all agents are allowed to form constant conjectures, and at least one agent is allowed to form linear conjectures. Then E has an infinite set of price-ratified market conjectural equilibria.*

Proof: Without loss of generality, let agent 1 be the agent with linear conjectures. A linear conjecture function C^1 may be decomposed into conjectures for individual goods $C^1(z^1) = (C_1^1(z_1^1), \ldots, C_n^1(z_n^1))$, where $C_j^1(z_j^1) = \alpha_j + \beta_j z_j^1$. Agent 1 is therefore strategic, with an optimal excess demand expressible as a function of α and β.[11] Let agents $i \neq 1$ adopt constant conjectures of the form $C_j^i(z_j^i) = P_j$. In equilibrium, the markets must clear. For all j,

$$z_j^1(\alpha, \beta) + \sum_{i=2}^{n} z_j^i(P) = 0. \tag{10}$$

For price-ratified equilibrium, we also require that agent 1's price conjecture for all goods j be equivalent to the other agents' conjectures, $\alpha_j + \beta_j z_j^1 = P_j$. We define a function

$$F(P, (\alpha, \beta)) = \begin{bmatrix} z_j^1(\alpha, \beta) + \sum_{i=2}^n z_j^i(P) \\ P_j - \alpha_j - \beta_j z_j^1 \end{bmatrix},$$

where $j = 1, \ldots, m-1$. From the discussion above we have that $F(P, (\alpha, \beta)) = 0$ implies price-ratified market conjectural equilibrium. Since α, β, and P are each m-vectors with $m-1$ degrees of freedom, F represents the mapping $F : \Re^{m-1} \times \Re^{2(m-1)} \to \Re^{2(m-1)}$. The conditions on utility functions ensure that excess demand functions are continuous, and thus that F is continuously differentiable. The conditions also ensure the existence of a competitive equilibrium P^*, and therefore there is a point $(P^*, (P^*, 0))$ such that $F(P^*, (P^*, 0)) = 0$. Then by the Implicit Function Theorem (Spivak, 1965), there exists an open set $\mathcal{P}$ containing P^* and an open set $\mathcal{B}$ containing $(P^*, 0)$ such that for each $P \in \mathcal{P}$, there is a unique $g(P) \in \mathcal{B}$ such that $F(P, g(P)) = 0$. All of these points $(P, g(P))$ constitute market conjectural equilibria for E.   ∎

Note that the conditions of Theorem 3 are satisfied by our experimental setup of Section 5. In that situation, the initial β determined which of the infinite conjectural equilibria was reached. Adding more strategic learning agents (those that could express non-constant conjecture functions) can only add more potential equilibria.

6.2.   *Dynamics*

The dynamics of a multiagent market system are dictated by how each agent changes its conjecture function, C^i, as it observes the effects of its chosen z^i on the price vector P. The strategic learning process given by Equations (5) and (6) can be transformed into the following system of differential equations, assuming that we allow continuous adjustment. For all j,

$$\begin{aligned} \dot{\alpha}_j &= \gamma_1(P_j - \alpha_j - \beta_j z_j), \\ \dot{\beta}_j &= \gamma_2(P_j - \alpha_j - \beta_j z_j)/z_j. \end{aligned}$$

Note that all variables are functions of time. The z_j solve the strategic agent's optimization problem (7), thus each is a function of α and β.[12]

Since the market determines prices based on specified demands, we can usually express P_j as a function α and β as well. The exception is when Equation 10 has no solution, for example when the strategic agent demands resources that the competitive agents are not willing or able to supply at any price.[13] This can happen only when the strategic agent's conjecture is highly inaccurate—but this is not ruled out by the system dynamics. An alternative price-adjustment algorithm—one that does not require an exact market clearing at each stage—may not be as sensitive to this problem.

For cases where the market always clears, the system of differential equations can be rewritten as

$$\begin{aligned} \dot{\alpha}_j &= \gamma_1 f_j(\alpha, \beta) \\ \dot{\beta}_j &= \gamma_2 f_j(\alpha, \beta)/z_j(\alpha, \beta), \end{aligned}$$

where $f_j(\alpha, \beta) = P_j(\alpha, \beta) - \alpha_j - \beta_j z_j(\alpha, \beta)$.

The equilibrium $(\bar{\alpha}, \bar{\beta})$ of this system is the solution of the following equations:

$$f_j(\alpha, \beta) = 0, \quad j = 1, \ldots, m - 1.$$

Since there are $m - 1$ equations with $2(m - 1)$ unknowns, the equilibrium is not a single point but a continuous surface, expressed as $(\bar{\alpha}, \bar{\beta}(\bar{\alpha}))$, where $\bar{\alpha} \in \Re^{m-1}$.

Characterization of conditions under which this learning process converges to a stable equilibrium remains an open problem.

6.3. *Perfect Conjectures*

Our experiments demonstrate that a learning agent might be rendered better or worse off by behaving strategically rather than competitively. However, the ambiguity disappears if it has sufficient knowledge to make a perfect conjecture. In the case where all the other agents are effectively competitive, perfect conjectures correspond to perfect knowledge of the aggregate demand function faced by the agent.

THEOREM 4 *Let economy E satisfy conditions for existence of competitive equilibrium. Then knowledge of the aggregate excess demand function of the other agents is a sufficient condition for an agent to achieve utility at least as great as it could by behaving competitively.*

Proof: Let agent 1 be the strategic agent, and z^1 its excess demand. Suppose the strategic agent knows the aggregate excess demand function of the other agents, $z^{-1}(P)$. Agent 1 knows that in market equilibrium,

$$z^1 + z^{-1}(P) = 0. \tag{11}$$

Therefore, the choice set Γ for the strategic agent consists of all excess demand bundles that could make the markets clear:

$$\Gamma = \{-z^{-1}(P) : P \in \Re^m\}.$$

If agent 1 behaves competitively, then any outcome it obtains must be part of a competitive equilibrium at some prices P^*. But by the market clearing condition (11), such an outcome must be contained in the strategic choice set Γ. Therefore, by optimizing over Γ, the knowledgeable strategic agent can achieve utility at least as great as obtained through competitive behavior. ∎

Intuitively, if the agent makes a perfect conjecture, then it makes its choice based on the actual optimization problem it faces. Any other choice would either have lower (or equal) utility, or violate the budget constraint.

As we have seen, however, when a strategic agent has imperfect information of the aggregate excess demand—for instance, a linear approximation—it may actually perform worse than had it used the constant approximation of competitive behavior.

7. Related Work

There is a growing literature on learning in games, much of it concerned with conditions under which particular protocols converge to Nash equilibria. Numerous studies have investigated the behavior of simple learning policies such as Bayesian update or fictitious play, or selection schemes inspired by evolutionary models. Researchers typically explore repeated games (especially coordination games), and have tended to find some sort of convergence to coordinated, equilibrium, or near-equilibrium behavior (Claus & Boutilier, 1998, Gilboa & Akihiko, 1991, Shoham & Tennenholtz, 1997).

Economists studying bidding games (Boyle, 1985, Samples, 1985) have noticed that biased starting bid prices strongly influence final bids. More generally, researchers have observed that the results of learning or evolution in games are often path-dependent (Young, 1996), with selection among multiple equilibria varying according to initial or transient conditions.

Most models in the literature assume that agents observe the joint action, as well as the resulting state. Our framework allows unobservable actions, and in the market game studied in depth, agents can reconstruct only an aggregate of other agents' actions. Boutilier (Boutilier, 1996) also considers a model where only outcomes are observable, demonstrating how to adapt some of the methods for the observable-action case to this setting. Interestingly, he finds that in some circumstances, uncertainty about other agents' actions actually speeds up the convergence to equilibrium for simple coordination games.

The last five years has seen some study of learning methods for agents participating in simple exchange markets. (Cliff's recent contribution (Cliff, 1998) includes a substantial bibliography.) Some of this work directly compares the effectiveness of learning strategic policies with competitive strategies. Vidal and Durfee examine a particular model of agents exchanging information goods (Vidal & Durfee, 1998), and find that whether strategic learning is beneficial (or how much) is highly context-dependent. We provide further data distinguishing the cases in our recent experiments within a dynamic trading model (Hu & Wellman, 1998a).

Finally, Sandholm and Ygge (Sandholm & Ygge, 1997) investigate a general-equilibrium scenario very similar to ours. Like us, they find that strategic behavior can be counterproductive when agents have incorrect models. Moreover, their study quantifies the costs of acting strategically and competitively as a function of model error, confirming that competitive behavior is far less risky for a range of environment parameters.

8. Conclusion

The fact that learning an oversimplified (in our case, linear) model of the environment can lead to suboptimal performance is not very surprising. Perhaps less obvious is the observation that it often leads to results worse than remaining completely uninformed, and adopting an even more oversimplified (constant) model. Moreover, the situation seems to be exacerbated by the behavior of the agent itself, optimizing with respect to the incorrect model, and thus "self-fulfilling" the conjectural equilibrium.[14]

Future work may shed some light on the situations in which self-fulfilling bias can arise, and how it might be alleviated. Random restart of the learning process is one straightforward

approach, as is any other deviation from myopic optimization aimed at trading exploitation for exploration. One could also expand the space of models considered (e.g., admitting higher-order polynomials), although it is clear that extending the class of conjecture functions can only add to the possible equilibria.

Another way to handle self-fulfilling bias is to transform this problem into a more traditional problem of decision under uncertainty. Agents that form probabilistic expectations may be less prone to get trapped in point equilibria. However, there is certainly a possibility of non-optimal expectations equilibrium even in this expanded setting.

A simple lesson of this exercise is that attempting to be a little bit more sophisticated than the other agents can be a dangerous thing, especially if one's learning method is prone to systematic bias. From a social perspective (or that of a mechanism designer), the prospect of disadvantageous conjectural equilibria might be a desirable property—discouraging agents from engaging in costly counterspeculations and potentially counterproductive strategic behavior.

More generally, our investigation serves to illustrate the role of equilibrium concepts—and specifically the application of conjectural equilibrium—in the analysis of multiagent learning. The interaction among dynamically evolving conjectures is what distinguishes the multiagent problem from its single-agent counterpart, and is thus arguably the learning phenomenon most worthy of the attention of multiagent systems researchers.

Appendix

The Strategic Agent's Optimization Problem

The nonlinear budget constraint faced by our strategic agents presents a problem more complicated than that of the standard competitive consumer. The specific form of the constraint depends on the conjecture function; our results below apply to strategic agents with linear conjectures, and thus quadratic budget constraints.

THEOREM 5 *Let the consumption set include all nonnegative bundles (i.e., $X = \Re^m_+$), and let U be a continuous function on X. Then there exists a solution to the strategic agent's optimization problem (7):*

$$\max_{z \in Z} U(z + e) \text{ s.t. } (\alpha + \beta z) \cdot z \leq 0. \tag{A.1}$$

Proof: To establish the existence of an optimum, we apply Weierstrass's Maximum Theorem (Horst, Pardalos, & Thoai, 1995): if S is a nonempty compact set in $\Re^m$, and $f(x)$ is a continuous function on S, then $f(x)$ has at least one global optimum point in S.

By assumption, the objective function U is continuous on X, and therefore also on $Z = \{z | z + e \in X\}$. Let S be the constraint set specified by (A.1), that is

$$S = Z \cap \{z | (\alpha + \beta z) \cdot z \leq 0\}.$$

We need to prove that S is a nonempty compact set in $\Re^m$. S is nonempty, since $(0, \dots, 0) \in S$. To show that S is compact is equivalent to showing that S is bounded and closed. It is obvious that S is closed. We prove that S is bounded.

From the constraint (A.1),

$$\sum_j \left(\alpha_j z_j + \beta_j z_j^2 \right) \leq 0, \text{ which implies}$$

$$\sum_j \left(\beta_j (z_j + \frac{\alpha_j}{2\beta_j})^2 \right) \leq K,$$

where $K = \sum \beta_j \frac{\alpha_j^2}{4\beta_j^2}$. Let $\hat{\beta} = \min\{\beta_1, \ldots, \beta_m\}$.

$$\sum_j (z_j + \frac{\alpha_j}{2\beta_j})^2 \leq \frac{K}{\hat{\beta}}$$

$$\left| z_j + \frac{\alpha_j}{2\beta_j} \right| \leq \left(\frac{K}{\hat{\beta}} \right)^{\frac{1}{2}}$$

$$|z_j| \leq \left(\frac{K}{\hat{\beta}} \right)^{\frac{1}{2}} + \left| \frac{\alpha_j}{2\beta_j} \right|$$

Thus S is bounded. By Weierstrass's theorem, U has at least one global maximum in S. Therefore there exists a solution to the stated optimization problem. ∎

THEOREM 6 *Let U be a continuous, strictly concave function on $X = \Re_+^m$. Then the optimization problem defined by (A.1) has a unique solution.*

Proof: Given the strict concavity of the objective function, and the existence of a solution (Theorem 5), it suffices to show that the constraint set S is convex.

Let $z', z'' \in S$, and $z = \lambda z' + (1 - \lambda) z''$, where $\lambda \in [0, 1]$. We need to show that $z \in S$. Let

$$\Delta_1 = \sum_j \alpha_j z_j + \beta_j z_j^2$$

$$= \sum_j \alpha_j (\lambda z_j' + (1 - \lambda) z_j'') + \beta_j (\lambda z_j' + (1 - \lambda) z_j'')^2$$

$$\Delta_2 = \lambda \sum_j \left(\alpha_j z_j' + \beta_j (z_j')^2 \right) + (1 - \lambda) \sum_j \left(\alpha_j z_j'' + \beta_j (z_j'')^2 \right)$$

Since $z', z'' \in S$, we have

$$\sum_j \left(\alpha_j z_j' + \beta_j (z_j')^2 \right) \leq 0, \text{ and}$$

$$\sum_j \left(\alpha_j z_j'' + \beta_j (z_j'')^2 \right) \leq 0,$$

thus $\Delta_2 \leq 0$. Therefore,

$$\Delta_1 \leq \Delta_1 - \Delta_2$$
$$= -\lambda(1 - \lambda) \sum_j \beta_j (z_j' - z_j'')^2$$
$$\leq 0,$$

since $\beta_j \geq 0$ for all j. $\Delta_1 \leq 0$ implies $z \in S$. Thus we proved that S is a convex set. Therefore the solution is unique.   ∎

The logarithmic utility function used in our experiments (Section 5) satisfies the conditions above, and thus our agent's problem has a unique solution. We solve the problem numerically using Lagrangean techniques.

Acknowledgments

Notes

1. Where exactly one draws the line between a change in beliefs and a simple update (incorporation of observational evidence) is fundamentally a matter of definition, and often quite arbitrary. We take no position, except to argue that any study that purports to characterize a learning process must clearly define this line, as does the framework proposed here.
2. Elsewhere, following Vidal and Durfee (Vidal & Durfee, 1997, Vidal & Durfee, 1998), we have distinguished between 0-level learning agents, which form models of the effects of their own actions, and 1-level learning agents, which form models of other agents (as 0-level agents). Recursive application defines higher levels. The question of which hypothesis space to adopt for multiagent learning problems is an interesting current research issue. Our investigations to date suggest that the appropriate form of target model can be highly problem specific, depending on observations available, and relative sophistication of other agents (Hu & Wellman, 1998a). We formulate our conjectural equilibrium concept in 0-level terms, to which higher levels can be reduced.
3. A more sophisticated version of this model would have agents form probabilistic conjectures about the effects of actions, and act to maximize expected utility.
4. Investigations of multiagent learning within the Markov game framework brings state dynamics to the fore (Filar & Vrieze, 1997, Hu & Wellman, 1998b, Littman, 1994).
5. The relationship between basic computational resources and results of computation can be modeled explicitly by extending the exchange economy to include *production*. See our prior work for detailed examples of general-equilibrium models of computational problems (Mullen & Wellman, 1995, Wellman, 1993, Wellman, 1995).
6. It is possible to express somewhat more general sufficient conditions in terms of underlying preference orders, but the direct utility conditions are adequate for our purposes.
7. In the standard model, no exchanges are executed until the system reaches equilibrium. In so-called *non-tatonnement processes* (Takayama, 1985), agents can trade at any time, and so the endowment e is also a function of time. In either formulation, we still assume that agents are myopic, optimizing only with respect to the current time period.
8. Cobb-Douglas utility is a limiting case of the CES form (constant elasticity of substitution),

$$U(x_1, \ldots, x_m) = \left(\sum_j a_j x_j^\rho \right)^{\frac{1}{\rho}},$$

   with $\rho \to 0$ (Arrow, et al., 1961). CES is commonly used in general equilibrium modeling (Shoven & Whalley, 1992), including some of our prior work. We also performed experiments with CES agents ($\rho = \frac{1}{2}$, and $a_j = 1$ for all j), with results qualitatively similar to those reported for the logarithmic case.
9. For configurations with only competitive agents (whether adaptive or simple), the system converges to the unique competitive equilibrium regardless of initial expectations.

10. In both (8) and (9), violation of feasibility results in consumption of the agent's own endowment. Reasonable definitions differing in the "otherwise" condition are also conceivable.

11. Here we refer to the vectors $\alpha = (\alpha_1, \ldots, \alpha_m)$ and $\beta = (\beta_1, \ldots, \beta_m)$, since the excess demand for good j generally depends on conjectures about the prices for all goods.

12. For a proof that a unique solution exists, see the appendix.

13. For example, the strategic agent's demand could exceed total endowments. For our example case of uniformly weighted logarithmic (Cobb-Douglas) utility, any demand exceeding $(m-1)/m$ times the total endowment of the competitive agents for any good is infeasible.

14. Kephart et al. (Kephart, Hogg, & Huberman, 1989) describe another setting where sophisticated agents that try to anticipate the actions of others often make results worse for themselves. In this model, the sophisticated agents' downfall is their failure to account properly for simultaneous adaptation by the other agents.

References

K. J. Arrow, H. B. Chenery, B. S. Minhas, & R. M. Solow. (1961). Capital-labor substitution and economic efficiency. *Review of Economics and Statistics*, 43, 225–250.

Boutilier, C. (1996). Learning conventions in multiagent stochastic domains using likelihood estimates. *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, (pp. 106–114). Portland, OR.

Boyle, K. (1985). Starting point bias in contingent valuation bidding games. *Land Economics*, 61, 188–194.

Brandenburger, A. (1992). Knowledge and equilibrium in games. *Journal of Economic Perspectives*, 6(4), 83–101.

Cheng, J.Q. & Wellman, M.P. (1998). The WALRAS algorithm: A convergent distributed implementation of general equilibrium outcomes. *Computational Economics*, 12(1), 1–24.

Claus, C. & Boutilier, C. (1998). The dynamics of reinforcement learning in cooperative multiagent systems. *Proceedings of the National Conference on Artificial Intelligence*, (pp. 746–752). Madison, WI.

Cliff, D. (1998). Evolving parameter sets for adaptive trading agents in continuous double-auction markets. *Agents-98 Workshop on Artificial Societies and Computational Markets*, (pp. 38–47). Minneapolis, MN.

Filar, J. & Vrieze, K. (1997). *Competitive Markov Decision Processes*. Springer-Verlag.

Fudenberg, D. & Levine, D.K. (1993). Self-confirming equilibrium. *Econometrica*, 61, 523–545.

Fudenberg, D. & Tirole, J. (1991). *Game Theory*. MIT Press.

Gibbons, R. (1992). *Game Theory for Applied Economists*. Princeton University Press.

Gilboa, I. & Akihiko, M. (1991). Social stability and equilibrium. *Econometrica*, 59, 859–867.

Grefenstette, J.J., et al. (Eds.). (1996). *AAAI Spring Symposium on Adaptation, Coevolution, and Learning in Multiagent Systems*. AAAI Press.

Hahn, F.H. (1977). Exercises in conjectural equilibrium analysis. *Scandinavian Journal of Economics*, 79, 210–226.

Horst, R., Pardalos, P. & Thoai, N. (1995). *Introduction to Global Optimization*. Kluwer Academic Publishers.

Hu, J & Wellman, M.P. (1996). Self-fulfilling bias in multiagent learning. *Second International Conference on Multiagent Systems*, (pp. 118–125). Kyoto, Japan.

Hu, J & Wellman, M.P. (1998a). Online learning about other agents in a dynamic multiagent system. *Second International Conference on Autonomous Agents*, pages 239–246, Minneapolis.

Hu, J & Wellman, M.P. (1998b). Multiagent reinforcement learning: Theoretical framework and an algorithm. *Fifteenth International Conference on Machine Learning*, (pp. 242–250). Madison, WI.

Kaelbling, L.P., Littman, M.L. & Moore, A.W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4, 237–285.

Kephart, J.O., Hogg, T. & Huberman, B.A. (1989). Dynamics of computational ecosystems. *Physical Review A*, 40, 404–421.

Littman, M.L. (1994). Markov games as a framework for multi-agent reinforcement learning. *Eleventh International Conference on Machine Learning*, (pp. 157–163).

Milgrom, P & Roberts, J. (1991). Adaptive and sophisticated learning in normal form games. *Games and Economic Behavior*, 3, 82–100.

Mullen, T. & Wellman, M.P. (1995). A simple computational market for network information services. *First International Conference on Multiagent Systems*, (pp. 283–289). San Francisco, CA.

Negishi, T. (1962). The stability of a competitive economy: A survey article. *Econometrica*, 30, 635–669.

Ono, N. & Fukumoto, K. (1996). Multi-agent reinforcement learning: A modular approach. *Second International Conference on Multiagent Systems*, (pp. 252–258). Kyoto, Japan.

Russell, S. & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall.

Samples, K. (1985). A note on the existence of starting point bias in iterative bidding games. *Western Journal of Agricultural Economics*, 10, 32–40.

Sandholm, T. & Ygge, F. (1997). On the gains and losses of speculation in equilibrium markets. *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, (pp. 632–638). Nagoya, Japan.

Sen, S. (1996). IJCAI-95 Workshop on Adaptation and Learning in Multiagent Systems. *AI Magazine*, 17(1), 87–89.

Sen, S. (Ed.). (1997). *AAAI-97 Workshop on Multiagent Learning*. AAAI Press.

Shoham, Y. & Tennenholtz, M. (1997). On the emergence of social conventions: Modeling, analysis, and simulations. *Artificial Intelligence*, 94, 139–166.

Shoven, J.B. & Whalley, J. (1992). *Applying General Equilibrium*. Cambridge University Press.

Spivak, B. (1965). *Calculus on Manifolds*. Benjamin/Cummings.

Takayama, A. (1985). *Mathematical Economics*. Cambridge University Press.

Tan, M. (1993). Multi-agent reinforcement learning: Independent vs. cooperative agents. *Proceedings of the Tenth International Conference on Machine Learning*. Amherst, MA: Morgan Kaufmann.

Vidal, J.M. & Durfee, E.H. (1997). Agents learning about agents: A framework and analysis. In Sen (Sen, 1997).

Vidal, J.M. & Durfee, E.H. (1998). Learning nested agent models in an information economy. *Journal of Experimental and Theoretical Artificial Intelligence*, 10(3), 291–308.

Weiß, G. (1993). Learning to coordinate actions in multi-agent systems. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, (pp. 311–316).

Wellman, M.P. (1993). A market-oriented programming environment and its application to distributed multicommodity flow problems. *Journal of Artificial Intelligence Research*, 1, 1–23.

Wellman, M.P. (1995). A computational market model for distributed configuration design. *AI EDAM*, 9, 125–133.

Young, H.P. (1996). The economics of convention. *Journal of Economic Perspectives*, 10(2), 105–122.