On the Emergence of Social Conventions: modeling, analysis, and simulations *

Yoav Shoham

Robotics Laboratory
Department of Computer Science
Stanford University
Stanford, CA 94305, USA

Moshe Tennenholtz

Faculty of Industrial Engineering and Management
Technion – Israel Institute of Technology
Haifa 32000, Israel

Abstract

We define the notion of *social conventions* in a standard game-theoretic framework, and identify various criteria of consistency of such conventions with the principle of individual rationality. We then investigate the emergence of such conventions in a stochastic setting; we do so within a stylized framework currently popular in economic circles, namely that of *stochastic games*. This framework comes in several forms; in our setting agents interact with each other through a random process, and accumulate information about the system. As they do so, they continually reevaluate their current choice of strategy in light of the accumulated information. We introduce a simple and natural strategy-selection rule, called *highest cumulative reward* (HCR). We show a class of games in which HCR guarantees eventual convergence to a rationally acceptable social convention. Most importantly, we investigate the efficiency with which such social conventions are achieved. We give an analytic lower bound on this rate, and then present results about how HCR works out in practice. Specifically, we pick one of the most basic games, namely a basic *coordination* game (as defined by Lewis), and through extensive computer simulations determine not only the effect of applying HCR, but also the subtle effects of various system parameters, such as the amount of memory and the frequency of update performed by all agents.

*This work has been partially supported by AFOSR grant AF F49620-94-1-0090.

1 Introduction

In multi-agent systems, be they human societies or distributed computing systems, different agents (people in the one case, programs or processes in the other) aim to achieve different goals, and yet these agents must interact either directly by sharing information and services, or indirectly by sharing system resources. In such distributed systems it is crucial that the agents agree on certain rules, in order to decrease conflicts among them and promote cooperative behavior. Without such rules even the simplest goals might become unattainable by any of the agents, or at least not efficiently attainable (just imagine driving in the absence of traffic rules). These rules strike a balance between allowing agents sufficient freedom to achieve their goals, and restricting them so that they do not interfere too much with one another.

We have been investigating social rules as a design tool. Some of these rules are designed and agreed upon ahead of time (traffic laws are an example); in previous work [21, 25] we investigated some aspects of this off-line design of social laws. However, not all rules can be agreed upon in advance. This is either because the characteristics of the society are unknown, or because they change over time. In addition, the design of all rules in advance might be computationally hard. In such cases, it is often important that the society converge on a convention in a dynamic fashion. In human societies this is common; this is how (e.g., software) standards emerge long before they are enshrined in official regulations.

How do such conventions emerge? Roughly speaking, the process we aim to study is one in which individual agents occasionally interact with one another, and as a result gain some new information. Based on its personal accumulated information, each agent updates its behavior over time. The complexity of this process derives from its concurrent nature: As one agent adapts to the behavior of the agents it has encountered, these agents update their behavior in a similar fashion. This tends to result in complex system dynamics, reminiscent of those encountered in particle physics, population genetics, and other areas. Each of these areas has developed stylized settings in which to carry out the investigations; we ourselves will adopt the framework of stochastic games from the economics literature.

In general terms, we will be asking two types of question: (1) Under what conditions do conventions eventually emerge? and (2) How efficiently are these conventions achieved? As it turns out, our results on eventual convergence will be primarily analytic, whereas the results on efficiency include both analytic lower bounds and empirical results of extensive computer simulations.

Here is the structure of our article, explained at two levels of granularity: a brief, jargon-free description, followed by a more detailed description that appeals to game theoretic terminology.

The brief description of the article is as follows:

- We give a formal definition of social laws and conventions, which are essentially the restriction of choices available to agents.
- We identify those laws and conventions that might be deemed rational from an individual standpoint.
- We define a stochastic setting in which agents interact with one another and update their behavior as a result. We define a particular update rule, and show that in certain circumstances it is guaranteed to lead all agents to accept a rational social convention.
- We then investigate how fast such rational conventions might emerge. We first give an analytic lower bound, and then investigate the actual rate of convergence in a particular case through extensive computer simulations.

Here is a more detailed overview of the article, which makes reference to game theoretic terms. The reader unfamiliar with game theory should just skim the following, and perhaps refer back to it once all the terms have been defined in subsequent sections.

- We adopt without change the notions of *games, payoff matrices,* and *rationality* as *utility maximization.* We also make reference to the notions of *maximin values, Nash equilibria,* and *Pareto optimality.* We make no novel contribution in this part.
- Next we consider the possibility of limiting the agents to a subset of the original strategies of a given game, thus inducing a sub-game of

the original one. We call such a restriction a *social constraint*; if the restriction leaves only one strategy to each agent, it is called a *(social) convention*. Some social constraints are consistent with the principle of individual rationality, in the sense that it is rational for agents to accept those (assuming all others do as well). In fact, we identify several senses of "rational social behavior." Some constraints are not rational in any reasonable sense. Both rational and irrational types of constraints may be of interest from a design standpoint, but we will pay special attention to the former. As social constraints fall within the general area of *cooperative games* in economics, whatever contribution we make in this part takes the form of added concreteness, a somewhat new perspective, and the attendant new terminology.

- Classical results in game theory make strong assumptions; in particular, they rely on the game being common knowledge. Much recent work in economics is devoted to investigating more realistic models. One important strand of recent work in economics, which has been strongly inspired by models of population genetics (e.g., [12, 9]), tends to relax not only the assumption that the game is common knowledge, but sometimes even that the game is known at all. Specifically, a number of models have been proposed in which agents engage in some process of (typically pairwise) interactions through which they gain information about the system (specifically, about how well each of their strategies has fared so far, perhaps about the strategies used by other agents, and, if the game is not known in advance, about the game). The agents may then use that information to update their choice of strategy, and the process repeats. It is then sometimes possible to show that the system will converge to a particular global state as if the players in fact had complete information and were acting rationally. The models within economics vary widely on how agents accumulate information, and how they update their choice of strategy. One important model is that of *stochastic games* and the notion of *evolutionary stable strategies (ess's)*, where it is shown that under certain conditions the iterated process will converge to a Nash equilibrium. We make no novel contribution to this work as such.

The items below constitute the core of our article, and are all novel.

- We ask, in an analogous fashion, how desirable social conventions might emerge through a stochastic process. These social conventions are not necessarily Nash-Equilibria.[1] We adopt the framework of stochastic games mentioned above. However, that framework allows quite a few variants, and our particular setting has unique features (we will explain and motivate these features later). Most importantly, we define a simple and natural strategy-selection rule called *highest cumulative reward* (HCR). (Again, for the reader familiar with ess's, we remark that this rule replaces the *best response* rule.)

- We show a class of stochastic games in which the HCR rule is guaranteed to converge to a rational social convention.

- We then ask how fast such social conventions might be achieved; most of our article is in fact devoted to this last topic. We first give an analytic lower bound on how fast it can be expected to be reached given *any* strategy-selection rule (we use a coupon-collector-style argument). We then investigate how fast such conventions evolve in practice. We do so by picking the simple *coordination* game, as defined by Lewis [2] , and through extensive computer simulations determine not only the effect of applying HCR, but also the subtle effects of various systems parameters, such as the amount of memory and frequency of update performed by all agents.

2 Games, social laws, and conventions

In this section we lay out the static framework, starting with the standard game-theoretic notions, and overlaying those with the notions of social laws and conventions.

[1] Although in some cases they may be. Our study will differ from the related studies in Economics on various other dimensions as well.

[2] In fact, we choose the perhaps most simple coordination game.

2.1 Games

All definitions in this section are standard and, in fact, very basic, in game theory. We include this section to make the article self-contained for those not familiar with game theory, and also to be clear about just how much we are taking from game theory (although we will take a bit more when we get to stochastic games). We start by defining the standard notion of a (one-shot) *game*. Intuitively, a game involves a number of players, each of which has available to it a number of possible strategies.[3] Depending on the strategies selected by each agent, they each receive a certain payoff.[4] The payoffs of the different agents are in general independent of one another, and are captured in a *payoff matrix*. Formally:

Definition 1 [k-person game]: A k-person game is defined by a k-dimensional matrix M, the entries of which are each a k-long vector of real numbers.

Intuitively, each dimension of the matrix represents the possible actions of the k players of the game. The j'th element of the vector $M(i_1, i_2, \ldots, i_k)$ represents the feedback to the j'th player if the actions taken by all the players are $i_1, i_2, \ldots, i_k$, respectively.

In this article we will be concerned exclusively with *symmetric games*. Intuitively, in symmetric games all players have the same strategies available, and the feedback they get does not depend on their roles or identities. More precisely:

Definition 2 [symmetric game]: A payoff matrix M defines a symmetric game iff the following hold:

1. All dimensions of M are of equal length, l. (Intuitively: The agents all have the same strategies available.)
2. For all $i_1, \ldots, i_k$ ($1 \leq i_j \leq l$, where $1 \leq j \leq k$) and $1 \leq m, n \leq k$, if $i_m = i_n$ then the mth and nth elements of the vector $M(i_1, \ldots, i_k)$ are identical. (Intuitively: Two players who play identically get the same payoff.)

[3]Some work in AI uses the term 'action' rather than the term 'strategy'; we will use both terms interchangeably.

[4]Some work in AI uses the term 'reward' instead; again, we will use both terms.

3. If $(i_1, i_2, \ldots, i_k)$ is a permutation of $(j_1, j_2, \ldots, j_k)$ then the vectors $M(i_1, i_2, \ldots, i_k)$ and $M(j_1, j_2, \ldots, j_k)$ are the corresponding permutations of one another. (Intuitively: The payoff to the players does not depend on their roles in the game.)

In addition to the restriction to symmetric games, throughout most of the paper we will concentrate on 2-person-2-choice games (i.e., M will be a 2×2 matrix with k=2). In the remainder of the article, and unless specified otherwise, a game will be understood to be a symmetric 2-person-2-choice game, and thus will have a matrix of the following form:

$$M = \begin{pmatrix} x, x & u, v \\ v, u & y, y \end{pmatrix}$$

Here are two examples of games. These two games, which are well-known in the literature, capture the phenomena of coordination and cooperation, respectively. Intuitively, the first game describes a situation in which the goal is to reach homogeneity in the society; it is an instance of the class of coordination games as defined by Lewis in his study of conventions [18].

Example 1: [a coordination game]:

$$M = \begin{pmatrix} 1, \ \ 1 & -1, -1 \\ -1, -1 & 1, \ \ 1 \end{pmatrix}$$

The second game we will consider is an instance of the well known *prisoners dilemma* setting, of the sort studied, for example, by Axelrod [2]. This game is a basic game in the study of cooperation.

Example 2: [a cooperation game, aka prisoners' dilemma][5] :

$$M = \begin{pmatrix} 1, \ \ 1 & -3, \ \ 3 \\ 3, -3 & -2, -2 \end{pmatrix}$$

[5]The reason we prefer the term 'cooperation game' to 'prisoners' dilemma' is that in cooperative, or bargaining, situations, which are the sort that we will consider, there is no dilemma associated with the game.

(In the cooperation game we call the first strategy available to each player 'cooperate' (or 'c' for short), and the second 'defect' (or 'd').)

The general question asked is, given a game, what strategies might the various players select. The combination of strategies selected by all the agents is called their *joint strategy* (or *joint action*). A basic assumption of game theory, which underlies many of its famous theorems, is that individuals are rational in the sense of being utility maximizers; that is, they will pick strategies that guarantee them the highest payoff. A number of important notions arise as a result; here are three.

1. If an agent knows what game is being played, but cannot assume that the other players do (or alternatively that they are rational), he might consider taking those actions that guarantee him the highest minimal payoff, no matter what the other agents do. The amount of this payoff is called *maximin value*. An action that guarantees the maximin value is called a *maximin stratgey*.

2. If the game the agents play is common-knowledge then the maximin strategy may not be the best choice; the worst-case scenario for a given agent might be also a non-optimal case for the other agent(s), and therefore can be assumed not to arise. A more appropriate notion in such setting is that of a *Nash equilibrium;* this is any joint strategy that is stable in the sense that no single agent benefits from switching to another strategy if all others remain unchanged. Nash equilibrium is among the most influential notions in game theory.

3. Another influential notion is that of *Pareto optimality.* A joint action is Pareto optimal if there does not exist another joint action that increases the payoff to one agent without decreasing the payoff to another.

Example 1 (cont.): In the coordination game the maximin value obtained is -1; both strategies are maximin strategies. There are two Nash equilibria, namely the two joint strategies on the main diagonal, and in both the payoff to each player is 1. These Nash equilibria happen to also be the two Pareto-optimal joint strategies in the game.

Example 2 (cont.): In the cooperation game the (unique) maximin strategy is 'defect,' with a maximin value of -2; this is also the strategy that will be performed in the (unique) Nash equilibrium. Nevertheless, this Nash equilibrium is the only joint strategy that is *not* Pareto optimal.

2.2 Social laws and conventions

Notions such as Nash equilibria make sense in a competitive setting that is devoid of any central control. In such a setting one can reasonably argue, for example, that in the cooperation game it is irrational for an agent to do anything but defect (and hence the paradox, since the agents are better off if they both cooperate).

However, consider a setting in which there does exist some central authority, be it a government in a human society or a system administrator in an electronic society. In this case, the authority may step in and dictate constraints. In general, it may dictate any constraint at all, in a way that is independent of the individual payoffs. This is an interesting possibility from a design standpoint, since there may be design goals that are not reflected in the individual payoffs. Indeed, our own primary motivation lies in using social laws as a tool for designing effective distributed systems. Nevertheless, in this article we will concentrate on constraints that serve the goals of the individual agents. Specifically, we consider the following scenario. Each agent is presented with the opportunity to accept constraints on his actions, conditional on all other agents accepting similar constraints. The constraints will be imposed if and only if all agents accept them, and in that case compliance with the constraints is guaranteed by the central authority. The question is what sort of social constraints are rational for the agent to accept under these conditions.

Definition 3 [social law]: A *social law* is a restriction on the set of actions available to the agents. A game g and a social law sl induce a sub-game g_{sl} of g that is the restriction of g to actions that are not prohibited by sl.

We may now define criteria according to which a social law may or may not be deemed rational. The tool we have at our disposal consists of the various variables defined on games, such as the three already mentioned – the maximin value, the set of values of the various Nash equilibria, and the set of values of the Pareto-optimal strategies. For any such variable V, let

$V(g)$ denote the value of that variable in the game g.[6] At this point in the article we remain agnostic about the choice of game variables; we will be less vague about it when we discuss the evolution of social conventions.

Definition 4 [rational social law]: Let g be a game, V a game variable, and $<$ an ordering on the possible values of this variable. A social law sl is *rational* with respect to g and V if $V(g) < V(g_{sl})$.[7]

The reader should notice that rationality here does not imply optimality. We view the acceptance of a suggestion made by the designer as rational if it improves upon what could be obtained without such suggestion.[8]

Of special interest are social laws that restrict the agents' behavior to a particular action:

Definition 5 [social convention]: A social law that restricts the agents' behavior to one particular strategy is called a *(social) convention.*

In most of this paper we will be concerned with simple games where each agent has to decide from among two actions; hence, we will be mostly interested in social conventions.

Example 1 (cont.): In the coordination game, there are two rational social conventions with respect to the maximin value, namely restriction to the first strategy and restriction to the second.

Example 2 (cont.): In the cooperation game there is one social convention that is rational with respect to the maximin value, namely restriction to 'cooperate.'[9]

[6] Recall that in this article we are restricting the discussion to symmetric games, and so we need not worry about different players attaching different values to a game variable.

[7] In a case that the game variable refers to a set of elements (such as the set of Nash Equilibria) we take $<$ to be an ordering over sets. In the case of maximin, the meaning of $<$ is straightforward.

[8] This does not imply of course that we view an agent who accepts a suggestion which does not improve upon its situation as irrational. However, we are especially interested (given our interest in symmetric games) in social laws which enable the agents to improve upon their situation.

[9] By showing that cooperation is a rational convention we do not mean to imply that there are not other settings that sanction cooperation; see [2].

3 Stochastic games and emergent conventions

As was mentioned in the introduction, classical results in game theory make strong assumptions; in particular, they rely on the game being common knowledge. Much recent work in economics is devoted to investigating more realistic models. One important strand of recent work in economics, which has been strongly inspired by models of population genetics, tends to relax not only the assumption that the game is common knowledge, but sometimes even that the game is known at all. Specifically, a number of models have been proposed in which agents engage in some process of (typically pairwise) interactions through which they gain information about the system (specifically, about how well each of their strategies has fared so far, perhaps about the strategies used by other agents, and, if the game is not known in advance, about the game). The agents may then use that information to update their choice of strategy, and the process repeats. It is then sometimes possible to show that the system will converge to a particular global state as if the players in fact had complete information and were acting rationally.

The models within economics vary widely on how agents accumulate information, and how they update their choice of strategy. One important model is that of *stochastic games* and the notion of *evolutionary stable strategies (ess's)*, where it is shown that under certain conditions the iterated process will converge to a Nash equilibrium. Kandori *et al.* [12] show that by gathering detailed statistics about the relative success of different strategies in a symmetric game that is played stochastically, and adopting a rule which says the society moves in the direction of the more successful strategies, subject to certain mutations, the system converges to Nash equilibrium. In [13], Kandori and Rob extend some of the results of [12]; they use the best response update rule, where a player selects the strategy that is the best strategy assuming other agents keep using their strategies (which are assumed to be learned by stochastic interactions). The best response update rule is also adopted in the work of Gilboa and Matsui [9], as well as in additional related work [15]. The main feature of the above-mentioned work is characterized by the fact it uses a model of *global* interactions where any agent interacts stochastically with all other agents, until gathering almost full information either on the strategies adopted by other agents or on the success of various strategies. This is quite different from models of local interactions where the agents are assumed to interact only with certain neighbors and to update their behavior in a more frequent manner. A detailed discussion on global and local models of interactions appear in [15]. We will return to this point later in the end of section 3.1, when we discuss a novel aspect of the model we use.

In the above discussion we mentioned some results on the emergence of Nash equilibria. We are interested in obtaining similar results for social laws and conventions. After all, the process we described for adopting a social law (the one in which agents were presented with the opportunity to voluntarily give up some options) made the same strong assumptions as classical work in game theory; in particular, it relied on the game being known (though not necessarily commonly known), and on agents being rational. We now ask whether social conventions, and perhaps even rational ones, might emerge also without these strong presuppositions. Specifically, we ask whether they might emerge through a stochastic process similar to the framework of stochastic games mentioned above. However, that framework allows quite a few variants, and our particular setting has somewhat unique features.

3.1 From Static to Stochastic Games

Definition 5 [n-k-g stochastic social game]: An n-k-g stochastic social game consists of a set of n agents, a k-person game g, and an unbounded sequence of ordered tuples of k agents selected from a uniform distribution over the n given agents.[10]

Intuitively, a stochastic social game describes a process in which, repeatedly, random k agents meet and play the particular game. In each iteration the actions are selected by the agents who participate in the game in a synchronous fashion. When agent i is selected to play in the game g in one of the rounds of n-k-g, i must select an action from among the actions available for it in the game g. An important question is what freedom we have in defining the action-selection function (which we will also call the *update rule*). We adopt two principles in this regard:

[10]The uniform-distribution assumption is made to simplify the discussion, but it can be relaxed and the results in the paper can be generalized suitably.

Obliviousness: The selection function cannot be based on the identities of agents or the names of actions.

Locality: The selection function is purely a function of the agent's personal history; in particular, it is not a function of global system properties.

We capture these principles in the following definition:

Definition 6: A selection function is *local* if it is a function of the history of actions taken by the agent and the corresponding payoffs received. A selection function is *semi-local* if it is a function of the history of actions taken, the corresponding actions taken by the other agents encountered by the agent, and the corresponding payoffs. In both cases it is required that a permutation of the names of actions in the history lead to a corresponding permutation of the actions selected.

Notice that a local selection function obeys both the locality and the obliviousness principles. A semi-local selection function is oblivious, but allows to refer to the actions performed by the agents encountered. The intuition behind the above principles is perhaps more important that its mathematical definition. We are interested in emergent rational social conventions in cases in which we cannot anticipate in advance the games that will be played. For example, if we know that the coordination problem will be that of deciding whether to drive on the left of the road or on the right, we can very well use the names 'left' and 'right' in the update rule; in particular, we can admit the trivial update rule which has all agents drive on the right immediately. Instead, the type of coordination problem we are concerned with is better typified by the following example. Consider a collection of manufacturing robots that have been operating at a plant for five years, at which time a new collection of parts arrive that must be assembled. The assembly requires using one of two available attachment widgets, which were introduced three years ago (and hence were unknown to the designer of the robots five years ago). Either of the widgets will do, but if two robots use different ones then they incur the high cost of conversion when it is time for them to mate their respective parts. Our goal is that the robots learn to use the same kind of widget. The point to emphasize about this example is that five years ago the designer could have stated rules of the general form "if in the future you have several choices, each of which has been tried this many times and has yielded this much payoff, then next time make the following choice"; the designer could not, however, have referred to the specific choices of widget, since those were only invented two years later.

This explains why we do not want the update rules to rely on *action* names. The prohibition on using agent identities in the rules (e.g., "if you see Robot 17 use a widget of a certain type then do the same, but if you see Robot 5 do it then never mind") is similarly motivated by the dynamic nature of the society; agents drop in and out of the society, denying the designer the ability to anticipate membership in advance. We definitely acknowledge that it is often useful to single out certain agents (such as Head Robot), and have them be treated in a special manner. We are very interested in the role of agents with special identities (and in particular in the role of organization structure [29]), but even with those it is still the case in a rich setting most of the agents will not be distinguishable in this fashion. In this article we investigate the emergence of successful joint actions only in such 'faceless masses,' and completely ignore the role of personal identities.

The above discussion concentrated mainly on the obliviousness requirement. Finally we need to motivate the requirement of locality. Interestingly, this requirement is not met by most dynamic systems put forward in similar settings. In particular, the work in economics discussed above, assumes that agents participate in sufficiently many interactions so as to have reliable global statistics about the system. This bias has its roots in the biological framework which inspired the economic model, and in particular in the global fitness function encountered in population genetics [1]. This global character of the update rule is even more blatant in the area of mathematical sociology [30], and in the work on computational ecologies [10]. It is not our claim that global information is never available to an individual in a society; counter-examples abound. However, it is clear that much of individual decision making is made in the absence of this global information, and our aim is to home in on this element. We will return to this topic when we compare our setting to dynamic system models in other fields, and, in particular, economics.

3.2 The Highest Cumulative Reward rule

We are now ready to start investigating useful action-selection rules. In [24] we reported on preliminary results of experiments with a number of such rules. Here we will concentrate on one particular local update rule, called *Highest Cumulative Reward.* There are a few reasons we concentrate on this rule. First, it is a very natural one. Second, past experiments have shown it to be particularly effective in stochastic settings. Finally, we will see that, despite its simplicity, this rule gives rise to nontrivial phenomena. (In the following definition, recall that in this article games are by default 2-person-2-choice games.)

Definition 7 [HCR]: According to the *Highest Cumulative Reward* update rule (or HCR), an agent switches to a new action iff the total payoff obtained from that action in the latest m iterations is greater than the payoff obtained from the currently-chosen action in the same time period.

The parameter m in the above definition denotes a finite bound, but the bound may vary. As we mentioned, HCR is a simple and natural update rule. It is, however, clearly not the only such rule. In particular, it would be natural to consider update rules that use a weighted accumulation of feedback rather than simple accumulation. Indeed, we have experimented with such rules as well. However, the results obtained, both analytic and experimental, were not qualitatively different from those obtained for HCR, and hence we stick with the simpler rule. A more detailed discussion of other update rules can be found in [24].

Clearly, HCR is a local update rule. (For the reader familiar with the relevant literature in economics, we remark that HCR stands in contrast to the *best response* update rule, in which the agent applies its best response to the set of strategies adopted by, essentially, all other agents.) We now would like to understand how HCR affects the emergence of rational social conventions, and, in particular, its effects on the evolution of coordination and cooperation. In fact, we are able to show a result that applies to a somewhat broader class of games, which include the coordination and cooperation games. We call these games *social agreement games.*

Definition 8:

A *social agreement game* is a symmetric game g with matrix

$$M = \begin{pmatrix} x, x & u, v \\ v, u & y, y \end{pmatrix}$$

in which $x, y, u, v \neq 0$, either $x > 0$ or $y > 0$ and either $u < 0$ or $v < 0$; if both $x > 0$ and $y > 0$ then $x = y$.[11]

It is easy to see that the cooperation game and the coordination game are both social agreement games.

The theorems below that refer to HCR assume that the parameter (memory bound) m is much larger than the entries in the payoff matrix of the game. We also assume that $m \geq n \geq 4$, and that the payoffs in g have finite decimal representation. With these assumptions, we have:

Theorem 1:[12] Given an n-2-g stochastic social agreement game, placing no constraints on the initial choices of action by all agents, and assuming that all agents employ the HCR rule, the following holds:

- For every $\epsilon > 0$ there exists a bounded number $\mathcal{M}$ such that if the system runs for $\mathcal{M}$ iterations then the probability that a social convention will be reached is greater than $1 - \epsilon$.
- Once the convention is reached, it will never be left.
- If a social convention is reached then it guarantees to the agent a payoff which is no less than the maximin value that was initially guaranteed.
- Furthermore, if a social convention exists for g that is rational with respect to the maximin value, then the social convention reached will be rational wrt maximin.

[11] It will be perhaps a bit jarring to some readers to see a formulation that depends on notions of 'positive' and 'negative' rewards, and thus one that does not allow a constant offset of all numbers. It is debatable whether the notions of 'positive' and 'negative' rewards are defensible; we believe that at the very least they are not trivially dismissed. Furthermore, even if one wished to do away with an objective notion of zero, one could perhaps synthesize one dynamically based (for example) on average payoffs encountered so far (related ideas appear, for example, in [23]). However, this discussion is beyond the scope of our paper.

[12] Proofs appear in the Appendix.

The above theorem shows that stable conventions can emerge using the a purely local update rule. In addition, it discusses also the evolution of stable conventions which are not Nash-Equilibria. In particular, our results show that using a purely local update rule a rational stable convention (w.r.t maximin) would emerge in the coordination and cooperation games:

Corollary 1: The HCR update rule guarantees eventual emergence of coordination and of cooperation, that is, rational conventions in the respective games.

3.3 The efficiency of evolution: a lower bound

The above results shed light on the eventual emergence of social behavior, but they say nothing about the efficiency with which this behavior is attained; the remainder of this article is devoted to this question. Our study of the efficiency of convention evolution will refer to the number of iterations required for obtaining a desired behavior. This measure of efficiency is different from the one which has been studied in models of stochastic global interactions. The measure of efficiency in that work has been the number of interaction periods which is required to reach a Nash-equilibrium. Each such interaction period consists of a huge number of iterations (in our terminolgy) where the agents gather information about each other. We start by presenting a general lower bound on the efficiency of convention evolution. This will be obtained by the following definition and theorem.

Definition 9:

Let g be a social agreement game. Consider iteration t of an n-2-g stochastic social game, and the $n \cdot (n-1)$ games (possible agent interactions) that might be played at that iteration. Define $X_n(t)$ to be a random variable that contains the number of games that might be played in iteration t and that result in a payoff for a player that is less than the one obtained by a rational social convention. Let $T(n)$ be a function that associates with each n a number (of iterations). Given a local update rule R, and some distribution on the initial actions of the agents, we will say that R *guarantees* the emergence of a rational social convention after $T(n)$ iterations, if $E[X_n(T(n))]$ converges to 0.

Roughly speaking, we measure how far the system is from reaching a rational social convention. We would like this distance to be as close to 0 as possible in a minimal number of iterations.

Theorem 2:

Let g be a social agreement game, and let R be a local update rule[13]. Assume there is some non-zero constant probability for starting with any particular action by any particular agent. If R guarantees the emergence of a rational social convention in the related n-2-g games in $T(n)$ iterations, then $T(n) = \Omega(n \cdot log(n))$.

4 The evolution of coordination: experimental results

At this point we seem to be converging on an understanding of the dynamics brought about by HCR; at least for social agreement games, we have a guarantee of eventual emergence to a rational social convention (if one exists), as well as a cautionary lower bound on how fast we can expect to arrive at such a happy occasion. It would be natural to expect that subsequent investigations would provide finer and finer lower and upper bounds, increasing our understanding of HCR.

Unfortunately, this has not been our experience. What we found instead was that rather specific properties of the particular games being played flavor the dynamics so strongly that it appears extremely difficult to arrive at general results at the level of the update function. We arrived at this conclusion through extensive computer simulations, which yielded results that not only had not been anticipated, but in fact have not yet been (fully) explained mathematically even after the fact.

Let us illustrate the point with the two games highlighted above, the coordination and cooperation games. Both are instances of social agreement games, and hence subject to the upper and lower bounds presented in the previous section, and yet the practical experience with the two has been

[13]Similar results hold for semi-local rules.

radically different. In the case of the coordination game, the HCR rule not only led to the emergence of convention, but it did so at a rate that approaches the theoretical lower bound. In contrast, in the cooperation game the HCR rule proved to be very inefficient, rendering it useless for most practical applications.

In the remainder of this article we restrict our attention to the coordination game, and explore various aspects of the efficiency with which coordination evolves.[14] Unless stated otherwise, when we refer to convention evolution, we will refer to the emergence of rational social convention in an n-2-g stochastic social game, where g is the coordination game. More specifically, when we say that a set of agents reached a convention, we mean that the agents in that set adopt the same strategy. All of our discussion and results remain valid when we replace the constant 1 in the coordination game, by any other constant $x > 0$. In this section we take the default value of m (in the definition of HCR) to be greater than the number of iterations (i.e., agents refer to their full history); we will be explicit when we depart from this default.

Unless stated otherwise, the experimental results appearing in this section refer to experiments with 100 agents starting with random initial strategies. Each experiment consists of many trials, each of which consists of a run of the stochastic game for a given number of iterations.

4.1 The effect of update frequency

The first parameter and modification we consider concern update frequency. In the previous section we assumed that each agent updates[15] its behavior at each iteration. What happens if agents update their behavior less frequently? This condition might be imposed by internal limitations of the system, or alternatively might be selected voluntarily to impose greater stability on the system.

A plausible a priori intuition about the effect of delaying the application of the update function might be as follows. If one does not delay at all,

[14] The efficiency of cooperation evolution is discussed further in [26].

[15] By 'update' we mean the application of the update function; the result need not be a change in action.

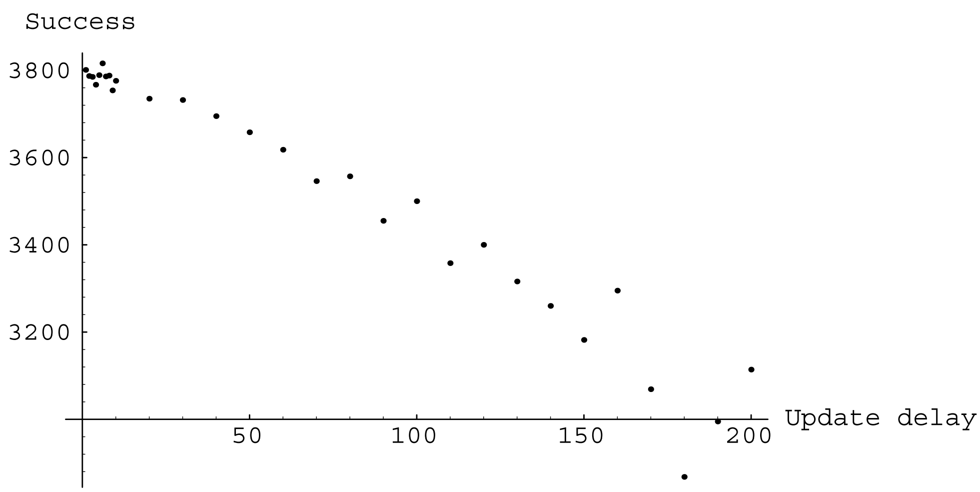

Figure 1: The effects of update frequency

agents may react on the basis of insufficient information, leading to a lot of thrashing in the system. If one delays too much, agents may be preventing from updating even when appropriate. So perhaps there is some optimal, middle-of-the-road course of action.

Particularly, in this setting this intuition isn't born out. We found that when the frequency of update decreases, then the efficiency of convention evolution decreases. Our results are illustrated by Figure 1. In this figure, the x coordinate describes the distance between iterations in which update is performed, while the y coordinate describes the number of trials from among 4000 trials of 1600 iterations each in which more than 95% of the agents reached a convention.

4.2 The effect of memory restarts

We investigated the effects of memory size on the efficiency of convention evolution. We consider two forms of limited memory; one is treated in this section, and the other one will be treated in Section 4.4. One type of limited memory is a memory that is restarted from time to time. When the memory is restarted, the agents' current strategies (the ones they will now start with) are not forgotten, but previous history is. This might be in particular inter-

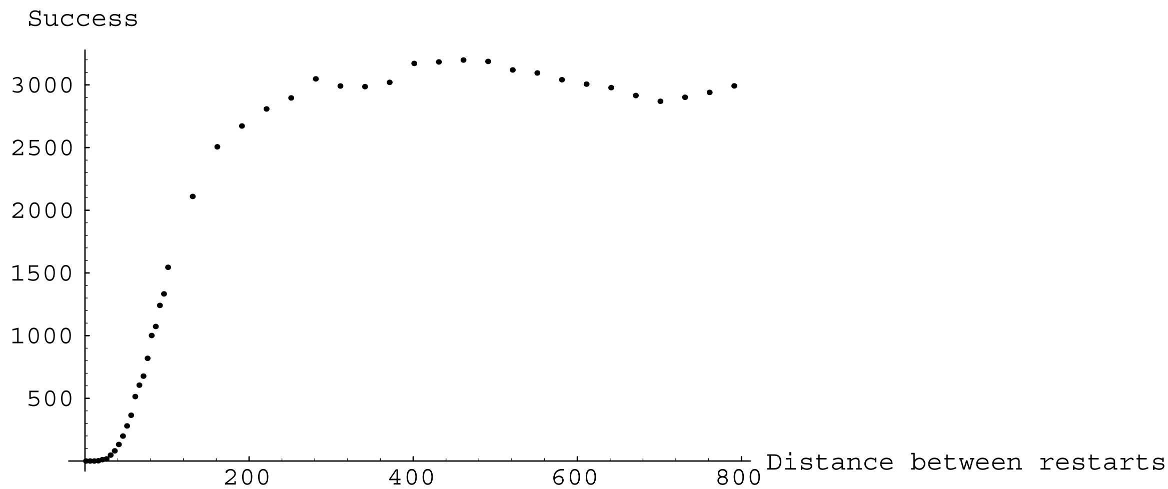

Figure 2: The effects of memory restarts

esting in systems which stop operating for a short while from time to time. For example, a society might be interested in a particular coordination only in some periods of the year, where agents are assumed to forget what they have exactly seen in the previous periods although they still remember their current (latest) strategy. We investigated the efficiency of convention evolution as a function of the frequency of memory restarts. We found that when the distance between iterations where the memory is restarted decreases, then the efficiency of convention evolution decreases. This is illustrated in Figure 2. The x coordinate of this graph corresponds to the distance between iterations where the memory is restarted. The y coordinate describes the number of trials from among 4000 trials of 800 iterations each, in which more than 85% of the agents reached a convention.

The reader may be tempted to treat this as an 'obvious' result; however, full memory is not always an advantage. Sections 4.3 and 4.4 will provide some examples; here is another example. We ran an experiment in which agents restarted their memory *always and only* after changing their strategy. In that case the evolution of convention was even more efficient than in the case of full memory; in 3298 from among 4000 trials of 800 iterations each, more than 85% of the agents reached a convention (while with complete information this was true of only 3010 of the trials.) We will explain why full memory is not always an advantage in the following sections.

4.3 Co-varying memory size and update frequency

We have so far varied update frequency and memory independently; we now show that these two parameters interact. Consider the results from section 4.1, where we showed that the rate of convention evolution is a monotonic increasing function of update frequency. We now show that decreasing memory blocks the degradation of convergence with the decrease in update frequency. Specifically, in this experiment we adopted the memory-restart model, and varied together the memory-restart frequency and the update frequency; that is, at the end of each window each agent updated its choice according to HCR for that window. The general result we obtained is that when update becomes infrequent (there is a long delay between strategy updates), then it is better to restart the memory from time to time than to rely on the whole memory. Our results are illustrated in Figure 3. The x coordinate of this figure corresponds to the update frequency, which is equal to the number of iterations between consecutive memory restarts. That is, in this case, we had a single interval which served both as the update frequency and the memory restart frequency. The y coordinate corresponds to the number of trials from among 4000 trials of 1600 iterations each in which 95% of the agents reached a convention. It is illuminating to compare Figure 3 to Figure 1 (where full memory is assumed); when the update frequency drops below about 100 iterations, it becomes better to use the statistics of only the last window than to rely on the entire history.

The rationale of the above result may be explained as follows. When agents have update delays they start relying on unreliable old information. By restarting its memory the agent succeeds in getting rid of some of this unreliable information.

One of the implications of the above result, from a design perspective, is that in systems where there are update delays the designer may wish to tell the agents to restart their memory from time to time in order to speedup the evolution of conventions. In the next section we will see that when there are no update delays even a more concrete kind of advice/result can be supplied/obtained.

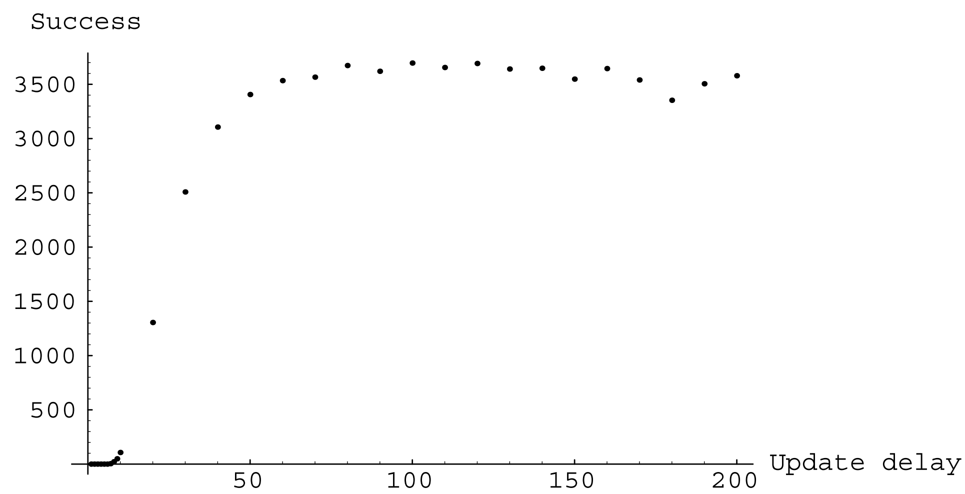

Figure 3: The case in which update frequency = memory restart frequency

4.4 Limited memory windows

A more continuous form of limited memory is one in which each agent at each time keeps a limited window into its past experience, and bases the HCR rule on only that window. We have considered two forms of windows, one in which an agent remembers the last m iterations in which it participated in a meeting, and another in which the agent remembers the last m iterations, regardless of whether it participated in a meeting in those.

Our results of these two experiments are illustrated in Figures 4 and 5, respectively. In both of these figures the x coordinate describes the size of the memory window, and the y coordinate corresponds to the number of trials from among 4000 trials of 800 iterations each, in which more than 85% of the agents reached a convention. Note that, somewhat surprisingly, in both cases it pays to forget, though some minimal memory is essential (in the first case this minimum is in fact equal to 2 iterations, and therefore this can be seen more easily in the second case).

The rationale of this result is that the old history of the agents is less adequate than the relatively new information, and as a result it may be better not to rely on old information as part of the data a decision refers to. On the other had, too short memory may not enable the agents enough sampling of what is going on in the system, and may lead to inefficient behavior.

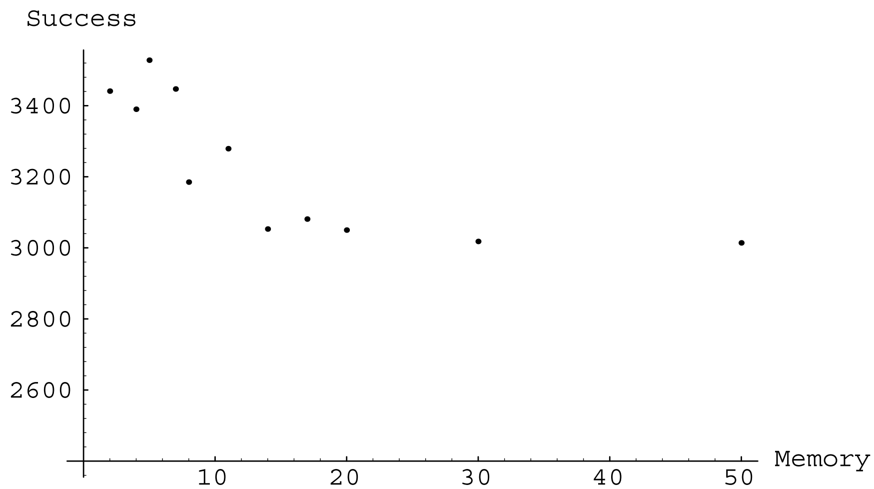

Figure 4: Limited Memory (latest observations)

A good choice of the memory window while applying HCR will give us in fact an update rule whose behavior is close to optimal. The case in which the memory size is between $2n$ to $3n$ (where n is the number of agents) gives us the above-mentioned close to optimal behavior, which is in fact a speed of convergence of $O(n \cdot log(n))$. More specifically, given that there are n agents who adopt HCR with a memory window $3n$ (where this number refers to the overall number of iterations, as in Figure 5), we observed that all of the agents reach a convention after less than $3n \cdot log(n)$ iterations (when we vary the number of agents.) The optimality stems from the above fact and from Theorem 2. The important point is that HCR with an appropriate limited memory window can be supplied to the agents as an update rule that will enable an efficient convention evolution in a system where there are no update delays.

One implication of this result is that it enables the designer to supply the agents with a concrete useful update rule which will enable conventions to evolve rapidly when there are no update delays. This may be of course most useful in situations where conventions are essential but can not be determined in advance.

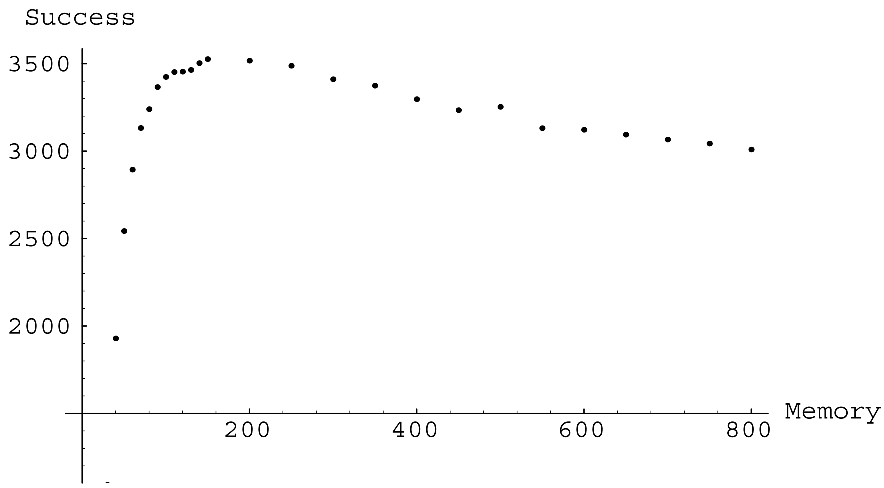

Figure 5: Limited Memory (latest iterations)

4.5 Further discussion of HCR

The previous sections have discussed several results about the efficiency of convention evolution. Our measure of efficiency has been the number of agents which adopt a convention after a given number of iterations. In particular, our graphs show the number of trials in which, after a particular number of iterations, the number of agents who adopt a similar (most popular) strategy is greater than a particular threshold. Our qualitative results do not change when different thresholds and numbers of iterations are used.

In addition to the above, one may be interested also in the dynamics of HCR for fixed assignments of the parameters. As it turns out, the dynamics are quite simple for any selection of the parameters. The number of agents who adopt the more popular strategy may have little fluctuations in the beginning of the process; then, this number increases until a convention is obtained; the speed in which this number increases, decreases along time. The explanation of these phenomena is as follows. The fluctuation appears mostly when the numbers of agents adopting different strategies are equally divided, and it is simply a result of the random selection of agents. The fact that the increase in the number of conforming agents is more modest towards the end of the process is explained by the fact that it takes time to a non-conforming agent to be selected by the process. We illustrate this

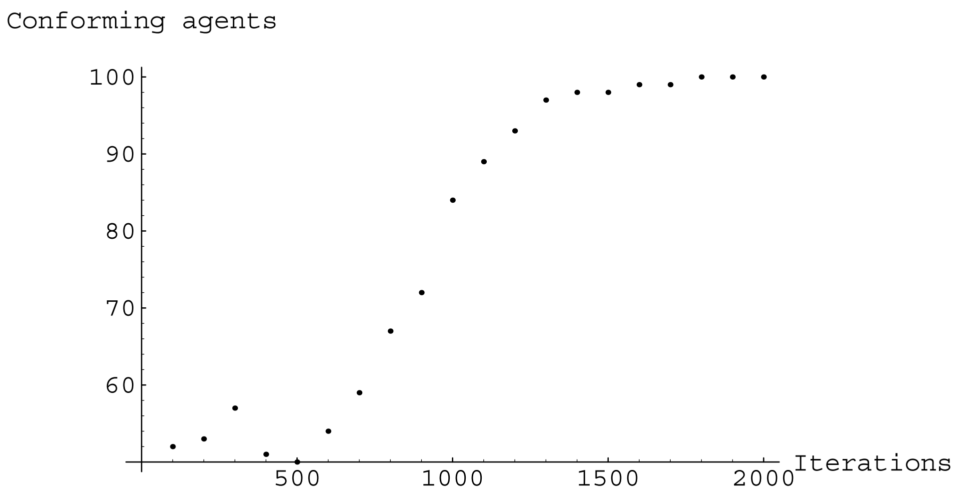

Figure 6: The shape of convention evolution

by Figure 6 , in which HCR is used by an agent with memory of 3000 (i.e., greater than the maximal number of iterations), and with no update delays. We consider the case where the strategies are equally divided among the agents in the beginning of the process. The x coordinate corresponds to the number of iterations, and the y coordinate corresponds to the number of agents conforming to the most popular strategy in that point.

4.6 More complicated decisions

The coordination game captures a situation where a selection among a pair of rational social conventions has to be made. This can also be considered as a selection of an option from among two possible options, without an a-priori agreement about which option should be chosen. What happens if the agents have to agree on an option from among more than two available options, that is, on something more complicated than a bit? How does the number of options (potential conventions) affect the efficiency of convention evolution?

In order to answer the above question we use the following observation: whenever an agent performs a particular strategy and gets a particular feedback in a 2-person-2-choice coordination game, it can interpret it as an observation of the strategy used by the agent it encountered. For example, if the agent performs strategy a and gets a feedback of 1, we can say that the agent observed that another agent used the strategy a as well. Having the above interpretation for the feedback, and assuming we restrict ourselves to quasi-local update rules only, we can define:

Definition 10: The External Majority (EM) update rule is an update rule which says: Adopt strategy i if so far it was observed in other agents more often than any other strategy and remain with your current strategy in a case no other strategy has been observed in other agents more often than it.

We can show:

Lemma 1: EM coincides with HCR in an n-2-g stochastic social game, where g is the coordination game.

Given the above lemma, HCR and EM are isomorphic in the context of 2-person-2-choice coordination games. Hence, although EM and HCR do not coincide when there are more than 2 choices in the coordination game, a natural extension of our study would be to discuss EM in the context of 2-person-s-choice coordination games.

Notice that the coordination game makes perfect sense, and it is of major interest, when agents are able to observe the behavior of agents they encounter. Moreover, even if the agents know the payoff matrix of the game, as well as are able to observe the behavior of agents they encounter, but do not have agreement on the names of strategies (i.e., the designer can not just tell them which strategy they should adopt) we still get a most interesting and fundamental problem. This is due to the full symmetry we have here. For example, the example we presented in Section 3 will still be valid in this case, as well as many other examples in the study of coordination [18]. In the sequel we will therefore assume a restriction only to quasi-local update rules, where the agents can observe the behavior of agents they encounter.

We would like to mention that by allowing quasi-local update rules, some of the power of our setting is lost. We will still be interested in local adaptation of the agents, in the sense that they may update their behavior at each iteration (and not in periods, in each of which the agent learns the strategies of the other agents, as in case of the best response update rule discussed in the economics literature), but the update rules may now refer to the strategies executed by the other agents in the past.

As mentioned, we would like to discuss the case in which the number of potential conventions is greater than 2:

Definition 11:

An *extended coordination game* is a symmetric 2-person-s-choice game, where the payoff for both agents is $x > 0$ if and only if they perform similar actions, and it is $-x$ otherwise.

Our general results are as follows. What we find is that adding more potential conventions decreases the efficiency of convention evolution in a less than logarithmic fashion. In addition we find that the absolute amount of success in convention evolution decreases in less than logarithmic fashion: For the number of successes of convention evolution to decrease by factor of 2, we need to increase the number of potential conventions by a factor of more than 4; for them to decrease by a factor of 3 we need to increase the number of potential conventions by a factor of more than 8.[16] Intuitively speaking, our results point to the following encouraging fact: the efficiency of convention evolution is not affected too badly by an increase in the number of potential conventions.

Some specific results are illustrated in Figure 7. The x coordinate describes on a logarithmic scale the number of potential conventions, while the y coordinate describes the number of successful trials (more than 85% reached a convention) from among 4000 trials of 800 iterations each.

The message of this result from a designer's perspective is that, by supplying an appropriate rule, the emergence of useful conventions is not hopeless also for complex systems where the number of potential conventions is more than two. Naturally, the emergence of more complex kind of conventions (e.g., where the convention itself is some structured strategy) may be an interesting subject for future research.

[16]We have verified these basic results also in the case of limited memory.

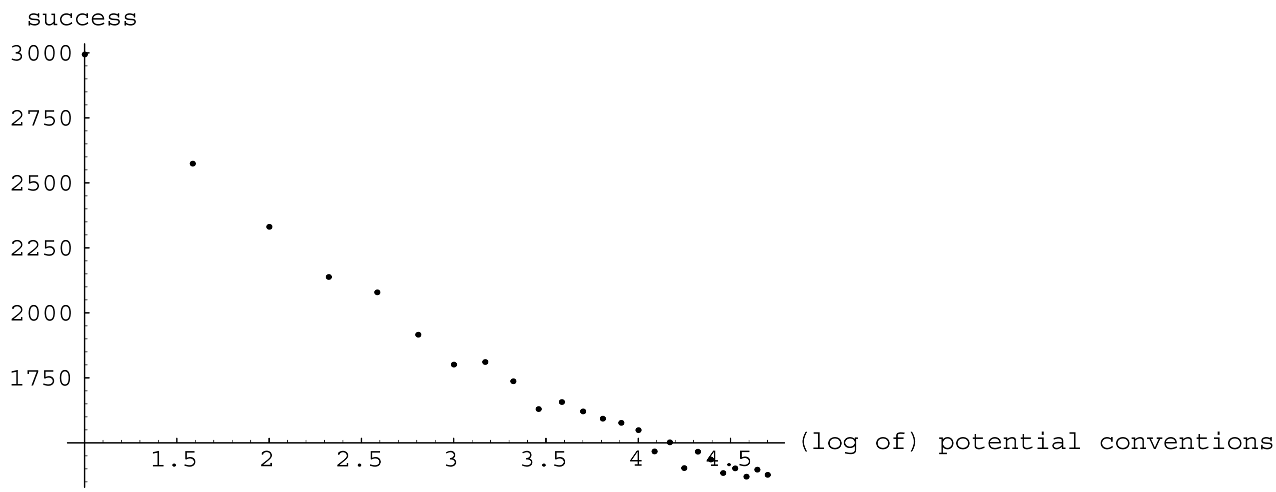

Figure 7: The effects of the number of potential conventions

5 Discussion and Related work

Several lines of research are related to our work. These include work in population genetics, statistical mechanics, computational ecologies, quantitative sociology, machine learning, and mathematical economics.

Recent work in *mathematical economics* is the one most related to our work, and was discussed in some detail in the previous sections. We would like to re-emphasize some of the major differences between our study and the related work in mathematical economics. The model we use is a model of global interaction [15] which has been borrowed from the related literature, but our adaptation process is local[17] ; the agent is not assumed to update its behavior after some period of interactions in which it is assumed to gather statistics about the other agents, but to obey a local adaptation rule where it is able to update its behavior at each point of the matching process. One other point is that in our study the agent need not be aware of the strategies taken by the other agents. In addition, our study concentrates on issues such as the efficiency of convention evolution. We also note that the efficiency of convergence in the above-mentioned related economic settings [7] refers to the

[17]The reader should be careful not to confuse at this point global interaction with global adaptation.

number of periods of interaction among the agents, where each period consists of a gathering of statistical information about the other agents. This is of course quite different from the type of efficiency studied in this paper. We also concentrate on the effects of various basic parameters of the adaptive scheme on the emergent behavior of the system. We study the effects of update delays, memory restrictions, and other parameters. Although memory plays a role in part of the related work in economics [31], the effects of this parameter (as well as of the other parameters discussed in this paper) on the efficiency of convention evolution has not been discussed. The final comparative point concerns the *type* of solutions we are interested in. In particular, some of our results refer to the emergence of social conventions that are not Nash equilibria.

Although work in Mathematical Economics is the most related to our work, the discussion in this paper would not be complete without at least a brief description of the work carried out in other related fields. Each of these works involve a setting with multiple elements (whether they are called particles, individuals, cells, or agents), which repeatedly undergo relatively simple local changes. The questions usually asked center around interesting global system properties that emerge over time out of these local changes, such as equilibria and phase transitions.

It is often tempting to try to carry over lessons from one setting to another. Indeed, some of these areas were inspired by one another; for example, work in quantitative sociology was inspired by work in statistical mechanics, and work in economics was inspired by work in population genetics. However, these inspirations have tended to be in spirit rather than in detail; the actual dynamic systems in the various areas are for the most part quite different, and also very sensitive in the sense that even small changes in them result in quite different system dynamics. This has also been our experience with our own framework; initially we had hoped to borrow results from other areas, but our framework then turned out to be sufficiently different from any of the others so as to make such borrowing impossible, or at least very difficult. We are still very much interested in understanding technical connections with these other areas and our own work, anticipating cross influences, but at this point all we will do is briefly describe work in these other areas.

Statistical mechanics models are a powerful tool for explaining a variety of phenomena in physics. An important family of models in that area goes under the general name of the *Ising model* [14]. In a typical Ising model we have a set of spins, each of which can be in -1/1 state, and which are organized into some fixed spatial arrangement (such as a one-dimensional sequence or a two-dimensional grid). At each point in time the system is in some configuration (that is, the spins each have a particular value), and this system has a certain measure energy, or entropy. The energy has a component representing local interactions among the spins, and a component (that is sometimes omitted) representing the effect of some global magnetic field. The interaction among spins is usually limited to neighboring spins; a typical formula for the energy of the system will include the sum of all multiples $x \cdot y$, where x and y are the values of neighboring spins. In terms of this energy a probability distribution is defined over the space of all configurations, which determines the likelihood of the system actually being in any particular configuration. This probability distribution has strong independence properties; the probability of a particular spin having a particular value is sensitive only to the values of the neighboring spins, and is independent of the values of spins that are spatially removed from it.

The Ising model has proved useful for the investigations of various physical properties such as spontaneous magnetization (that is, a majority of spins ending up with the same value), but it is also abstract enough to have motivated other applications. For example, in some work within *quantitative sociology* [30] the spins were interpreted as 'opinions,' the energy between individual spins as 'tension among individuals with differing opinions,' and the orientation of the magnetic field as 'the opinion of the government.'

As described, the Ising model does not provide a dynamical system, in the sense that it does not provide (e.g. differential) equations that describe the evolution of the system over time; what it does instead is define stable (i.e., low energy) states of the system. This is also true for the more elaborate framework of spin-glass. However, both have been augmented to include a dynamical system. These dynamic models have also found applications in other fields. One is quantitative sociology, where the models have been used to predict opinion shifts over time within large populations [30]. Statistical mechanics also provided the inspiration to *computational ecology* [10]; this work is based on the idea that the existence of many agents in an advanced computerized framework creates a "computational ecology." A computational framework, similar in its spirit to quantitative sociology, is developed and analyzed using the tools of statistical mechanics. The notions used there are 'strategies' of individual agents, the utility of having identical strategies ('cooperation') as well as its disadvantage, due to resource conflicts ('competition'). A precise continuous framework is built, which allows several predictions on the behavior of those "computational ecologies" (such as chaotic behavior in some situations).

These multi-agent frameworks borrow a powerful tool from statistical mechanics, but as a result they have a heavily 'non-local' or 'non-mechanical' flavor; the dynamics speak about how certain global statistics change over time (such as the average number of cooperating agents), rather than about how an individual agent changes its local state on the basis of its current local state and/or history. This of course is quite unlike our own framework, where the dynamics of the atomic changes are the basis for change, and any statistical properties are derived from these.

Work in *population genetics* (e.g., [1],[3],[19]) and work in *artificial life* inspired by it (e.g., [20],[16]) is closer to ours in this sense. Here we have a set of individuals, each belonging to one of several types. The system evolves in 'generations'; in each generation the individual evolves in a way that is defined by its type and the environment (which includes the other agents), and at the end of the generation the 'fitness' of each agent is computed, applying some given fitness function. The probability that an individual will survive into the next generation is proportional to its fitness. Additionally, usually between generations a process of 'recombination' takes place, in which some pairs of individuals in the population (the 'parents') may combine to produce a new individual ('the offspring'), whose type will in general be defined by, but different from, the types of the parents.

This setting is thus more local, or mechanical, than the frameworks discussed earlier; the activity of each agent within a generation and the reproductive process have a transition-oriented, automata-like flavor. An important global component remains, however, namely the fitness function. The fitness function, which represents unspecified external selection forces, is computed on the basis of global system properties, and is applied to all agents equally. For example, if we consider the earlier 'opinion' example, a typical definition of the fitness of an individual with an opinion in a population would be the proportion of individuals in the population having the same opinion. This global element turns out to have a strong influence on the dynamics of the system. (This is perhaps the deepest difference between the population genetics setting and our own; but of course may other differences exist, including the notion of an accumulated history, limited memory, and our particular stochastic process of encounters.)

The model used in population genetics has strongly influenced work in *mathematical economics*; this is especially true of work published in recent years. This recent work in mathematical economics is the closest in spirit to our work, and therefore we discussed it in detail previously.

The way the agents update their behavior in our setting, has some similarity with learning rules used in the reinforcement learning literature (e.g., [11]), and the fact that we have a multi-agent system where agents behave based on local feedback has some similarity with work on learning automata [22]. Nevertheless, as we mentioned, our work borrows a framework of agent interaction which is common in the recent mathematical economics literature inspired by theoretical biology. This makes our study much different from existing studies in reinforcement learning. Moreover, our objective, the study of emergent rational social conventions, is different as well.

Our work is clearly relevant to previous work on Artificial Social systems which we have mentioned before, as well as to theories of social commitments [17, 5] and social reasoning [27]. The research reported in this paper is concerned however with dynamic systems where social behavior is an emergent property of the system. The emergence of social reasoning and complex social commitments are beyond the scope of our current work, and may be a subject for further research. Another topic which may be relevant for future research is concerned with the concept of negotiations, which have been widely discussed in the distributed and decentralized AI literature [4, 6]. One can consider the situation where agents may have a limited negotiation ability which they may use when encountering each other in the stochastic setting;[18] based on these interactions the agents may learn for example about available strategies they have not considered in the past. Finally, since our work in concerned with the emergence of social conventions, it is related to work on cooperative games and mechanism design ([8], [32]) where cooperative solutions to multi-agent interactions are devised. However, since our work is concerned with the emergence of such cooperative solutions, it becomes more

[18]Some of the results in [24] are concerned with a limited form of such extension.

related to the above-mentioned work on complex dynamic systems.

6 Summary

We used the framework of stochastic social games in order to investigate the emergence of rational social conventions and the efficiency of that process. In particular we concentrated on the emergence of rational social conventions in a most basic type of coordination problem, and supplied results on the emergence of conventions and its efficiency for a class of games.

Besides the novelty of our work, which we have previously discussed, we believe that it also creates a bridge between work in economics and work in machine learning. We borrow from the economics literature the model for stochastic interactions among agents, which is a most popular and dominant model for agent interactions. On the other hand we borrow ideas of reinforcement learning [28] from the AI literature in order to capture the fact that agents use local update rules to update their strategies. As a result, we get a combined framework where local update rules are used to update an agent behavior in a model of global interaction. Since both the local updates ([28],[11],[22]) and the model of global interactions (see [15] for a survey) are taken to be of major importance to the corresponding communities, we believe that the introduction of our framework may lead to additional and fruitful cross-fertilization.

References

[1] L. Altenberg and M. W. Feldman. Selection, Generalized Transmission, and the Evolution of Modifier Genes. I. The reduction principle. *Genetics*, pages 559–572, November 1987.

[2] R. Axelrod. *The Evolution of Cooperation.* New York: Basic Books, 1984.

[3] A. Bergman and M. W. Feldman. More on Selection for and against Recombination. *Theoretical Population Biology*, 38(1):68–92, 1990.

[4] A. H. Bond and L. Gasser. *Readings in Distributed Artificial Intelligence.* Ablex Publishing Corporation, 1988.

[5] C. Castelfranchi. Commitments: From individual intentions to groups and organizations. In *1st International Conference on Multi-Agent Systems*, pages 41–48, 1995.

[6] Y. Demazeau and J.P. Muller. *Decentralized AI.* North Holland/Elsevier, 1990.

[7] G. Ellison. Learning, local interaction, and coordination. *Econometrica*, 61(5):1047–1071, 1993.

[8] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, 1991.

[9] I. Gilboa and A. Matsui. Social stability and equilibrium. *Econometrica*, 59(3):859–867, 1991.

[10] Bernardo A. Huberman and Tad Hogg. The Behavior of Computational Ecologies. In Bernardo A. Huberman, editor, *The Ecology of Computation.* Elsevier Science, 1988.

[11] L. Kaelbling. *Learning in Embedded Systems.* MIT Press, 1993.

[12] M. Kandori, G. Mailath, and R. Rob. Learning, Mutation and Long Equilibria in Games. Mimeo. University of Pennsylvania, 1991.

[13] M. Kandori and R. Rob. Evolution of Equilibria in the Long Run: A General Theory and Applications. Mimeo. University of Pennsylvania, 1991.

[14] R. Kinderman and S. L. Snell. *Markov Random Fields and their Applications.* American Mathematical Society, 1980.

[15] A. P. Kirman. Economies with Interacting Agents. SFI working paper, 94-05–030, 1994.

[16] J. Koza. Genetic Evolution and Co-Evolution of Computer Programs. In C.G. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, editors, *Artificial Life II.* Addison-Wesley, 1992.

[17] H. Levesque, P.R. Cohen, and J. H. Nunes. On acting together. In *Proc. of AAAI-90*, 1990.

[18] David Lewis. *Convention, A Philosophical Study.* Harvard University Press, 1969.

[19] U. Liberman and M. W. Feldman. A General Reduction Principle for Genetic Modifiers of Recombination. *Theoretical Population Biology*, 30(3):341–370, 1986.

[20] K. Lindgren. Evolutionary Phenomena in Simple Dynamics. In C.G. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, editors, *Artificial Life II.* Addison-Wesley, 1992.

[21] Y. Moses and M. Tennenholtz. On Computational Aspects of Artificial Social Systems. In *the Proceedings of DAI-92*, 1992.

[22] K. Narendra and M. A. L. Thathachar. *Learning Automata: An Introduction.* Prentice Hall, 1989.

[23] A. Schwartz. A Reinforcement Learning Method for Maximizing Undiscounted Rewards. In *Proceedings of the 10th International Conference on Machine Learning*, 1993.

[24] Y. Shoham and M. Tennenholtz. Emergent Conventions in Multi-Agent Systems: initial experimental results and observations. In *Proc. of the 3rd International Conference on Principles of Knowledge Representation and Reasoning*, pages 225–231, 1992.

[25] Y. Shoham and M. Tennenholtz. On the Synthesis of Useful Social Laws for Artificial Agent Societies. In *Proc. of AAAI-92*, pages 276–281, 1992.

[26] Y. Shoham and M. Tennenholtz. Co-Learning and the Evolution of Social Activity. Technical Report STAN-CS-TR-94-1511, Dept. of Computer Science, Stanford University, 1994.

[27] J. M. Sichman and Y. Demazeau. Exploiting social reasoning to deal with agency level inconsistency. In *1st International Conference on Multi-Agent Systems*, pages 352–359, 1995.

[28] R.S. Sutton. Special issue on reinforcement learning. *Machine Learning*, 8(3–4), 1992.

[29] M. Tennenholtz. On Computational Social Laws for Dynamic Non-Homogeneous Social Structurs. To appear in JETAI, 1994.

[30] W. Weidlich and G. Haag. *Concepts and Models of a Quantitative Sociology; The Dynamics of Interacting Populations.* Springer-Verlag, 1983.

[31] H. P. Young. The evolution of conventions. *Econometrica*, 61(1):57–84, 1993.

[32] G. Zlotkin and J. S. Rosenschein. A Domain Theory for Task Oriented Negotiation. In *Proc. 13th International Joint Conference on Artificial Intelligence*, pages 416–422, 1993.

Appendix: Proofs of Theorems

Proof of Theorem 1:

Recall that the payoff matrix of a social agreement game has the structure

$$M = \begin{pmatrix} x,x & u,v \\ v,u & y,y \end{pmatrix}$$

in which either $x > 0$ or $y > 0$, either $u < 0$ or $v < 0$, and if both $x > 0$ and $y > 0$ then $x = y$.

We prove the theorem by case analysis. We can assume without loss of generality that $x > 0$. Notice that the cooperation game is a special case of the case in which $y < 0, u < 0, v > 0$, and the coordination game is a special case of the case in which $y > 0, u < 0, v < 0$. We will provide the proof for these two cases; proofs for the other cases can be obtained in a similar fashion.

Consider the case where $y > 0, u < 0, v < 0$. In this case a rational social convention will restrict the behavior of all agents to a similar strategy. First, observe that there always exists a pair of agents with identical strategies. Then, notice that the following process can be generated with a probability $p = \frac{1}{f(n)}$ and leads to a rational social convention (all agents will adopt the same strategy) in $g(n)$ iterations, where both $f(n)$ and $g(n)$ are bounded by an exponent of the form n^s where s is polynomial in m (the memory size) and n. The process is defined as follows: a pair of agents (i, j) with the same strategy is selected and meet each other until all of the rest of the agents forget their past. Afterwards, i meets a member $x \neq j$, and then meets j. The last step continues in a loop where at each time i meets a new x until it meets all the members in the society. It is easy to see that this process will bring to a rational social convention (all agents will adopt the same strategy). As a result, if the system runs for $\mathcal{M} = k \cdot g(n) \cdot f(n)$ iterations then the probability that a rational social convention will not be reached (not all of the agents will adopt the same strategy) is at most e^{-k}. Taking $k > -log(\epsilon)$ yields the desired result.

Consider the case where $y < 0, u < 0, v > 0$. In the sequel we will refer to an agent who adopts the strategy c as a "cooperative" agent and to an agent who adopts d as a "non-cooperative" agent. In this case a rational social convention will restrict all of the agents to be cooperative. The structure of the proof is as the structure of the proof regarding the case where $y > 0, u < 0, v < 0$, but the basic process will now change. This process will now at first guarantee that there will be at least two cooperative agents. In order to guarantee this, the process will include in its beginning a procedure of creating a pair of cooperative agents (if no such pair exists). This procedure selects two non-cooperative agents and two additional agents, and let the latter pair meet until the former pair will forget its past. Afterwards the process selects the former (non-cooperative) agents to participate in a meeting. This will create a pair of cooperative agents. In a second stage this pair of cooperative agents will meet until the other agents will forget their past, and then pairs of non-cooperative agents will meet sequentially. This will create a society where at most one agent is non-cooperative. In order to make this agent cooperative the process will end with the following procedure: the non-cooperative agent will meet a cooperative agent until it will become non-cooperative as well, and then a pair of cooperative agents will be selected and meet each other until the rest of the agents will forget their past. The process will end by an encounter in which the two non-cooperative agents meet each other.

The above process will take place with probability $p = \frac{1}{f(n)}$ and will guarantee that after $g(n)$ iterations all of the agents will become cooperative, where appropriate exponential bounds can be given for $f(n)$ and $g(n)$. Hence, the number $\mathcal{M}$ can be calculated as for the coordination game, and the desired result can be obtained.

The results for the other cases are determined similarly. In all of these cases all agents will eventually adopt a similar strategy s, where the payoff for the joint strategy (s, s) is greater than 0. This implies that social conventions will be reached, and that they will never be dropped (given the structure of HCR). Notice also that any rational social convention of the game g will restrict the behavior of agents to a particular strategy where the payoffs for both agents is positive. Hence, given that if both $x > 0$ and $y > 0$ we have that $x = y$, and given the structure of emerged conventions discussed above, we get that if a rational convention exists it will also emerge in the related process.

▮

Proof of Theorem 2:

Let $Y_n(i)$ be a random variable which contains the number of agents that did not participate in any iteration of n-2-g until iteration i. It is easy to see that $E[X_n(i)] \geq k \cdot E[Y_n(i)]$ for some constant $k > 0$ and for every n and i. In particular, $E[X_n(T(n))] \geq k \cdot E[Y_n(T(n))]$ for every n. Hence, it suffices to show that if $E[Y_n(T(n))]$ converges to 0 as a function of n, then $T(n)$ is at least of the order of $n \cdot log(n)$. The probability that a particular agent will not be chosen along $T(n) = (n-1) \cdot f(n)$ iterations is bounded by $(1 - \frac{1}{(n-1)})^{2 \cdot (n-1) \cdot f(n)}$ which converges to $e^{-2f(n)}$. If $e^{-2f(n)} > \frac{1}{n}$ then we will get that $E[Y_n(T(n))] > 1$ and hence there is no convergence to 0. But, in order to have $e^{-2f(n)} \leq \frac{1}{n}$ we must have $f(n) \geq 0.5 \cdot log(n)$ (where we consider w.l.o.g the natural log). This gives us the desired lower bound.

∎

Proof of Lemma 1:

For ease of exposition let us denote the strategies by 0 and 1, and let c_i be the accumulated payoff for strategy i of a given agent j. Notice that c_i equals to the number of times that j met an agent which used i minus the number of times j met an agent which used $1 - i$, when its (j's) strategy was i. According to HCR, an agent chooses c_i if it is larger than c_{1-i}, but $c_0 - c_1 =$ (number of times you met 0 when you had 0 minus number of times you met 1 when you had 0) – (number of times you met 1 when you had 1 minus number of times you met 0 when you had 1), which equals to the number of times you met 0 minus the number of times you met 1. Hence, we get that the comparison between the accumulated payoffs coincide with the comparison between the number of times the different strategies were encountered in other agents. This gives us the desired result.

∎