11-2002

On Adaptive Emergence of Trust Behavior in the Game of Stag Hunt

Christina H Fang
University of Pennsylvania

Steven. O. Kimbrough
University of Pennsylvania

Stefano Pace

Annapurna Valluri
University of Pennsylvania

Zhiqiang Zheng
University of Pennsylvania

Recommended Citation

Fang, C., Kimbrough, S. O., Pace, S., Valluri, A., & Zheng, Z. (2002). On Adaptive Emergence of Trust Behavior in the Game of Stag Hunt. *Group Decision and Negotiation, 11* (6), 449-467. http://dx.doi.org/10.1023/A:1020639132471

On Adaptive Emergence of Trust Behavior in the Game of Stag Hunt

Abstract
We study the emergence of trust behavior at both the individual and the population levels. At the individual level, in contrast to prior research that views trust as fixed traits, we model the emergence of trust or cooperation as a result of trial and error learning by a computer algorithm borrowed from the field of artificial intelligence (Watkins 1989). We show that trust can indeed arise as a result of trial and error learning. Emergence of trust at the population level is modeled by a grid-world consisting of cells of individual agents, a technique known as spatialization in evolutionary game theory. We show that, under a wide range of assumptions, trusting individuals tend to take over the population and trust becomes a systematic property. At both individual and population levels, therefore, we argue that trust behaviors will often emerge as a result of learning.

Disciplines
Other Computer Sciences | Other Social and Behavioral Sciences | Theory and Algorithms

On Adaptive Emergence of Trust Behavior in the Game of Stag Hunt

Christina Fang*

Steven Orla Kimbrough* (kimbrough@wharton.upenn.edu)

Stefano Pace#

Annapurna Valluri

Eric Zheng

*** The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 USA.**

 # School of Management, University of Bocconi, Italy.

Abstract

In this paper, we adopt a learning perspective to shed light on the evolutionary origin of trust. It aims to answer three basic research questions. First, can individuals learn to behave cooperatively, regardless of their different dispositions? In other words, can trusting behaviors evolve over time as a result of trial and error learning? We model this possibility of process or emergent trust by a computer algorithm borrowed from the field of artificial intelligence [Watkins, 1989]. Second, is this learning effective? Specifically, we are interested in the performance implication of learning against different opponents adopting various strategies. For instance, will the agent learn to recognize an opponent playing Tit-for-Tat and behave cooperatively as a result? Third, if trust can indeed be learned, will it spread throughout a society of individuals and become a property of the entire system? This evolution of trust at the population level is modeled by a grid-world consisting of cells of individual agents, a technique known as spatialization in evolutionary game theory.

In contrast to prior research that views trust as hardwired, genetically transmitted traits, we show that trust can emerge as a result of trial and error learning. We provide an explicit mechanism of learning, which unravels the origins and developmental process of trust, both as a static norm and as an emergent property. Through learning, we demonstrate that trusting behaviors can evolve at both dyadic and societal levels.

Introduction

Allan Greenspan recently, and unexceptionally, underscored the critical nature of trust to our social and economic way of life, "[T]rust is at the root of any economic system based on mutually beneficial exchange…if a significant number of people violated the trust upon which our interactions are based, our economy would be swamped into immobility" [Greenspan, 1999]. Indeed, trust, or at least the principle of give and take is a pervasive element of social exchange.

Despite its centrality, trust is a concept fraught with ambiguities and even controversies. The concept means different things to researchers in different fields or even sub-fields. Indeed, it is not entirely clear that there is a single concept to be found. Trust is seen as embedded in the larger concept of social capital [Adler and Woo, 2000] by sociologists, while social psychologists interested in the same concept refer to 'emotional states and involuntary non-verbal behavior' as trust as well. Even organizational researchers find it hard to agree on a consistent set of definitions [Zaheer et al., 1996]. For example, more than 30 different definitions of trust have been found from a recent survey of literature [McKnight and Chervany, 1996].

Following the tradition of evolutionary game theory, we operationalize trust as the propensity to cooperate in the absence of other behavioral indicators [Macy, 1996]. While we acknowledge that cooperation may not be the result of trust, we believe that this operationalization is defensible because trust cannot be said to have existed if there is no manifested cooperation. In addition, equating trust with cooperation has been at the heart of a long established convention of both evolutionary and experimental game theories. As such, studying cooperative behavior represents a crude but important first step towards empirical verification. However, since we simply cannot refute the objection that the mechanisms we propose produce not trust, but merely its functional equivalent [Granovetter**,** 1985], we must refrain from making definitive statements about trust itself.

In contrast to competing conceptual analyses of trust, we are investigating three main research questions pertaining to trust itself. First, will individuals learn to behave cooperatively as a result of trial and error learning? We model this possibility of process or emergent trust by a computer algorithm borrowed from the field of artificial intelligence [Watkins, 1989]. Second, is this learning effective? We examine the performance implications of learning against different opponents adopting various strategies. For example, will the agent learn to recognize an opponent playing Tit-for-Tat and behave cooperatively as a result? Third, if trust can indeed be learned by individuals, will it spread throughout a society and become a property of the entire system? We model this evolution of trust at the population level by a grid-world consisting of cells of individual agents, a technique known as spatialization in evolutionary game theory.

By way of framing this paper, we report results from a series of experiments that *pertain* to the origin and emergence of trust behavior. Behavior, that is, which indicates or at least mimics the behavior of trusting individuals. We take a broadly game-theoretic perspective, in contrast, e.g., to a philosophical analysis or a social-psychological study of trust. We do not offer a definition of trust; nor do we think one is necessary. Instead, we appeal to established usage in the context of the well-known game called *stag hunt*. This game has long been of interest because when humans encounter it play is naturally described as evidencing trust or not, depending on what the players do. At the least, there is apparently trusting behavior and apparently non-trusting behavior. Our focus in this paper is on whether rather simple artificial agents will display 'trust behavior'—behavior corresponding to apparently trusting behavior of humans playing the same game—when they play stag hunt. To that end, we discuss next the game of stag hunt and what in it counts as trust behavior. From there, we move on to a discussion of our experiments with artificial agents in that game.

Learning to Trust in a Game of Stag Hunt

We begin with Prisoners' Dilemma, a game that highlights the stark conflict that may exist between what is best for each all concerned and 'rational' pursuit of individual ends. In the language of game theory, Prisoners' Dilemma is represented (in strategic form) by the payoff matrix in Table 1. The unique Nash equilibrium (marked by *) occurs when both actors end up defecting.

Insert Table 1 about here

The game is defined by the payoffs, not by the labels ('cooperate' and 'defect') on the players' options. The labels are conventional, and well-established, shorthand for describing what the options mean. We agree with the conventional usage: a row (column) player who chooses the top (left) option is behaving in an apparently cooperative or trusting manner; a row (column) player who chooses the bottom (right) option is behaving in an apparently uncooperative or defecting manner.

Using Iterated Prisoners' Dilemma (IPD, repeated plays of the game between fixed players), researchers have consistently found that cooperation, or trust, will evolve to be the norm under a broad range of conditions [Axelrod, 1980; Grim et al, 1999; Nowak and Sigmund, 1993]. This basic model has therefore become the 'e. coli of social psychology', and has been extensively applied in theoretical biology, economics, and sociology over the past thirty years. Researchers have been 'trying to shoehorn every example of cooperative behavior into this Prisoners' Dilemma since 1981", according to behavioral ecologist David Stephens [Morrell, 1995].

There has been extensive work on Prisoners' Dilemma [Macy and Skvoretz, 1996], yet game represents but one plausible model of social interaction in which the pursuit of individual self-interest will lead actors away from a mutually beneficial ('cooperative' or 'trusting') outcome. The results we report here are about a quite different and much less studied game involving the possibility of cooperation, stag hunt, attributed originally to Rousseau. Turning attention to this important game promises to enrich our understanding of trust in a novel problem context.

Stag hunt takes its name from a passage in Rousseau emphasizing that each individual involved in a collective hunt for a deer may abandon his post in pursuit of a rabbit adequate merely for his individual needs [Grim et al., 1999]:

> *When it came to tracking down a deer, everyone realized that he should remain dependably at his post, but if a hare happened to pass within reach of one of them, he undoubtedly would not have hesitated to run off after it and after catching his prey, he would have troubled himself little about causing his companions to lose theirs. (*Rousseau, Discourse on the Origin of Inequality, 1755*)*

Here, the study of trust is embedded in a context rather different from prisoners' dilemma. No individual is strong enough to subdue a stag by himself, but it takes only one hunter to catch a hare. Everyone prefers stag to hare, and hare to nothing at all (which is what a player will end up with if he remains in the hunt for stag and his counter-player runs off chasing hares). In this game, which is also called a trust dilemma in the literature, mutual cooperation takes on the highest value for each player; everything is fine as long as the other player does not defect. Cooperation against defection, however, remains far inferior to defection against either cooperation or defection. As such, there are two Nash Equilibria. One is the cooperative outcome of mutually staying in the hunt for stag. The other is the outcome mutual defection. See Table 2. This 'trust' game forms the context of our empirical investigations.

Insert Table 2 about here

Experiments in Individual-Level Learning

At the individual level, we model actors who are able to learn in repeated games and who may then learn (apparently) trusting behavior. We model, or simulate, this individual learning process by an algorithm known as Q-learning in artificial intelligence [Watkins, 1989; Sandholm and Crites, 1995]. Our simulation consists of two players playing the game of Stag Hunt iteratively for a specified number of times. To begin, we fix the strategies of one of the players [the Opponent] and examine how the other player [the Learner] adapts. It is our hypothesis that a cooperative (apparently trusting) outcome will be learned by the Learner if the Opponent also acts cooperatively.

Q-learning is widely used in artificial intelligence research [Sutton and Barto, 1998; Hu and Wellman, 1998; Littman, 1994]. It is part of the family of reinforcement learning algorithms, inspired by learning theory in psychology, in which the tendency to choose an action in a given state is strengthened if it produces results, weakened if unfavorable. This algorithm specifies a Q function, a function that depends on a pair of state and action variables, which keeps track of how good it is for the agent to perform a given action in a given state. There are only four possible outcomes in any play of a 2-player stag hunt game. Each player independently has 2 available actions: to cooperate [C] or to defect [D]. The Q function is therefore a 4 by 2 table with 8 cells as shown in Table 3. (Given each of the four outcomes of a game, the Q function has to decide whether to play [C] or [D] next.)

Insert Table 3 about here

As entries in the Q table store the value of taking a particular action given an observed state from the previous iteration, learning the relative magnitude of these values is key to effective adaptation. Such learning can occur through repeated exposure to a problem, when an agent explores iteratively the consequences of alternative courses of action. In particular, the value associated with each state-action pair (s, a) is updated using the following algorithm [Watkins, 1989]:

$$Q(S,A) \leftarrow (1 - \alpha)Q(S,A) + \alpha * (R + \gamma * Q(S', A'))$$

where α is the learning rate, γ the discount factor and (S', A') the next period state-action pair.[1]

As more games are played, the initially arbitrary beliefs in the Q function table are updated to reflect new pieces of information. In each game, the agent chooses probabilistically[2] the preferred action, observes the state of the game and the associated payoffs, and uses this information to update her beliefs about the value of taking the previous action. In this way, she learns over time that certain states are better than others. Since the cooperative outcome yields the maximum payoff, we might expect the Learner

1 For an introduction to the family of reinforcement learning models, see Reinforcement Learning by Sutton and Barto 1998.

2 The choice of action is guided by the so-called Softmax exploration method to prevent pre-mature locking in into local optima [See Sandholm and Crites, 1995 for details]. Essentially, the method ensures that all

will gradually choose to cooperate over time. Such emergence of trust or cooperation (behavior), however, is critically dependent on the strategies adopted by the Opponent. For instance, if the Opponent has a fixed strategy of always defecting, then any intelligent Learner should learn not to cooperate since doing so will always earn her 0 payoff. Therefore, to provide a more realistic picture of the dyadic dynamics, we need to account for the Opponent.

To simplify, we fix the strategies of the Opponent such that she behaves quite predictably. In general, any fixed strategies can be characterized by triples <I, c, d>, where I indicates the initial play, c the response to cooperation and d the response to defection on the Player's side [Grim et al., 1999]. The eight possible strategies can then be set out in a binary fashion as in Table 4[3]:

Insert Table 4 about here

For instance, the most commonly studied strategy Tit-for-Tat is described as <1, 1, 0> since an agent following such strategy will always cooperate in the first round, and later continue to cooperate unless the Opponent defects. Exhausting the 8 possibilities in this way allows us to model all available Opponent strategies. Therefore, we can study the efficacy of learning by pitting our Learner against each type of the Opponent.

Two modeling clarifications need to be mentioned here. First, there are other particularly relevant computational models of learning, notably genetic algorithms and neural networks. For instance, Macy & Skvoetz (1996) use genetic algorithms to model trust in an iterated Prisoners' Dilemma. These so-called evolutionary approaches differ from reinforcement learning in that they directly search the space of possible policies for one with a high probability of winning against the opponent. Learning therefore is said to occur off line through extensive training. Reinforcement learning, on the other hand, learns while interacting with the environment without conducing an explicit search over possible sequences of future states and actions [Sutton & Barto, 1998]. Second, learning

---

actions have a positive probability of being chosen in any given round. The degree of greediness in search is tuned by a temperature parameter τ, with smaller values of τ representing a greedier search process.

[3] We do not consider Pavlov, a recently proposed strategy by Nowak and Sigmund (1993). It cooperates on the first move and thereafter repeats its previous move if the opponent cooperated; otherwise it switches.

in a 2 by 2 game encompasses more than just a stationary game against nature since the behavior of the opponent may change over time. In multi-agent learning tasks in the context of a 2 by 2 game, one can either explicitly model the opponent or simply consider the opponent as a part of a non-stationary environment. While Hu and Wellman (1998) prove that Q-learning algorithms taking into account opponents do converge to Nash Equilibria under specific conditions, Banerjee, Mukherjee and Sen (2000) show that those who do not model opponents often out-perform those who do in the long run in their 2 by 2 game. Since the latter approach is far simpler and has been shown to work well in a game context, we have chosen not to model explicitly opponents' behavior.

Our results for individual learning are based on simulations of 500 plays each for a given type of Opponent. The Learner plays the iterated Stag Hunt game for 500 times with each type of the Opponent. The learning rate α is set at 0.999954 and set to be decaying at the same rate. The discount factor γ is set at 0.9. The softmax temperature τ is initially set at 30.0598, which reduces incrementally to 0.0598 such that eventually the search becomes almost greedy. We have chosen these parameter values after the values used by other papers [e.g., Sandholm & Crites, 1995].[4]

Simple Learning against a Fixed Opponent

The Q-learning mechanism ought to find reasonable strategies when playing against a fixed opponent. Therefore, as a validation test and for purposes of benchmarking, we begin with a series of experiments involving simple Q-learning by a Learner against a fixed Opponent in Iterated Stage Hunt. Table 7 summarizes both the frequency of occurrence across 5000 runs of 500 games **/* correct? */** of the four possible states in which the Learner plays against Opponents with fixed strategies and the associated payoffs for the players. The Opponents' strategies are indicated in the eight right-most column headings of the Table. The row headings (in the left-most column) indicate possible outcomes, e.g., CD means Learner cooperates and Opponent defects. **/* Is this correct? */** Indeed, after playing 500 games, cooperation emerges successfully in

[4] We use the payoffs shown in Table 2 as rewards in our simulation. In results not reported here, we alter the rewards associated with playing cooperation while the other player cooperate and find that the same qualitative patterns emerge. As such, our result is not sensitive to the particular payoff values we specify.

matches in which the Learner is pitted against relatively trusting Opponents such as those playing Tit for Tat or Quaker. The Learner quickly realizes that it is in the mutual best interest to play cooperatively and as such, the cooperative outcome [CC] becomes the dominant state which is observed around 87% of the time. **/* Not clear from the Table where this number comes from. Looks more like 100% to me (sok)*/** Trusting behavior, however, is not beneficial to the Learning if the Opponent tends to defect. Our Learner encounters difficulty in coping with Opponents who are nastier and trickier. When the Opponent always defects, the Learner manages to learn to acquire a slight aversion to cooperation as the mutually defecting state [DD] occurs a bit more frequently than [CD]. **/* ?? Looks like 100% DD ?? */** However, this result seems less than satisfactory as we should expect any intelligent Learner to learn to shun cooperation all together. Such is indeed the case when we reduce the degree of exploration in the action choice process.[5]

In the case of a tricky Opponent who 'maliciously' reverses her play (i.e. when she is some sort of a Doormat), our Learner is confused and rightfully so. It is not clear what a best reactive play against such an Opponent should be. Evidence of learning, however, is not totally absent. By looking at the payoffs obtained by the Learner and her Opponent, we find that even in such difficult cases, the Learner manages to out-perform her Opponent, despite the fact she fails to discover any consistent set of strategies.

Insert Table 7 about here

It is revealing to plot the evolution of various outcome states over time against Opponents playing three types of fixed strategies: Tit for Tat, Always Defect and Suspicious Defector.

Insert Figure 1 about here

---

[5] When we reduce the degree of exploration such that choice of action is made more sensitive to the relative magnitude of values in the Q table [by tuning down the exploration parameter Tao from 5 to 1], we observe that the percentage of [DD] occurring shoots up to about 96%. In other words, the Learner will learn to always defect.

In Figure 1, where the Opponent plays Tit for Tat, we see that the Learner eventually learns that to cooperate with the Opponent to attain the higher payoffs associated with the cooperative outcome. In Figure 2, where the Opponent always defects, the Learner also learns to avoid being taken advantage of. As such, we see that the loser outcome [CD] is almost completely avoided towards the end of the 500 episodes. At the same time, the Learner realizes that the best strategy against an opponent who always defects is simply to defect all the time. This knowledge is manifested by the fact that the Learner chooses in such a way that the outcome is the punishment [DD] 100% of the time.

Insert Figure 2 about here

In Figure 3, where a tricky Opponent such as a gullible doormat who plays opposite strategies (defect after seeing cooperation and cooperate after seeing defection), our Learner is confused and rightly so. She falls into the three outcomes [CC] [DC] and [DD] almost equally, indicating no marked preference for any one. This does not mean that there is no learning. Faced with a gullible doormat, our Learner clearly realizes that she should avoid the outcome [CD] since by cooperating, she knows that she will induce her Opponent to defect in the next round. As such, she does learn to shun the loser outcome [CD] almost entirely towards the end.

Insert Figure 3 about here

In a nutshell, we have shown that in cases where cooperation is in the best mutual interest of both players, our Learner has no trouble identifying and carrying out the cooperative solution when playing against Opponenent. Trust evolves purely as a result of individual learning as the agent adapts to the recurring game-theoretic problem. Such learning is quite effective, despite the fact that no model of the external environment is known at the outset. In exchanges in which the other player turns out to be ‘malicious’ or ‘un-trustworthy’, trust (behavior) fails to evolve and for good reasons. Finally, Figures 1-

3 are representative in that learning in these regimes becomes clear only near 500 generations of play.

Learning with more memory capacity

In the experiments of the previous section, both the Learning and the Opponent agents have a memory capacity of 1. They can only remember what happens in the last round, which in turn becomes the basis for their current action. It is reasonable to hypothesize that a 'smarter' agent, one with a larger memory capacity, will outperform her more mnemonically challenged counterparts. In particular, we would expect this intelligence to manifest itself especially in games in which the opponent's strategy is more complicated. To investigate this conjecture, we endowed agents with a memory capacity of 4 units – meaning that they can remember what transpires in the previous 4 episodes. We ran the agents again with two types of fixed strategy Opponents: Tit for Tat and Gullible Doormat. In the former case, we find that the number of episodes it takes for the pair to reach a cooperative equilibrium is much lower when the Learner can remember more. It takes only 180 episodes whereas it takes almost the entire 500 episodes if the agents have a memory capacity of 1. More memory clearly implies better performance. When the opponent is a gullible doormat, however, the situation is reversed. While with a memory capacity of 1 agents clearly learn to avoid becoming a benevolent loser, the agents with a memory capacity of 4 fail to learn anything. This result highlights a potential downside of being 'smarter' – sometimes remembering clearly other people's mistreatment can be dysfunctional if it results in less forgiving and therefore less cooperative behaviors.

Learning against another fellow learner

The experiments so far assume that there is only one party who is learning. The strategy of the Opponent is fixed. Results indicate that learners are not very effective against opponents who are pre-disposed to defect in one way or another. What will be the collective outcome if both parties can learn at the same time? On the one hand, if both agents can learn to realize the Pareto-enhancing feature of the cooperative outcome, games played by two learners will converge to the social optimal more quickly than before. On the other hand, since both agents are learning, one agent's action will alter the environment faced by the other. There is less stability simply because now both agents have to constantly adjust to the other, as in a real game.

We find that in about 84% of the runs the agents learn to achieve the mutually beneficial cooperative outcome. although they still fall into the inferior outcome [DD] around 17% of the times. **/* That's 101%. Please explain. */** Since both players are now learning, it is difficult to compare the results with any of the cases in the fixed Opponent analyses above. However, we can still conclude that two self-interested agents can often produce a Pareto-efficient outcome that maximizes the welfare of the pair together, as seen in Figure 4 below. **/* How do you get 84% and 17% from that figure? */**

Insert Figure 4 about here

Population level learning

The evolution of trust is quintessentially a problem beyond the simplicity of a 2 by 2 game. We now turn to learning to trust with multiple individuals and examine how trust can evolve into a systemic property. At the population level, we are primarily interested in whether and how trust can spread throughout the entire population. Specifically, we want to see how this evolution is influenced by 1) the initial percentage of trusting individuals and 2) the initial distribution of them.

To model the emergence of trust at the system level, we use a natural extension of the 2 by 2 game by embedding it in a spatial framework. In this set of experiments, we develop an 81*81 grid to represent the space of a population. Individual agents, represented by cells in a spatialized grid, behave entirely in terms of simple game-theoretic strategies and motivations specifiable in terms of simple matrices [Picker, 1997;

Grim et al., 1999]. We shall see how something resembling cooperation and even generosity can arise as a dominant pattern of interaction in a population of individuals primarily driven by self-interests. A given cell is occupied in the next round by a player of the type that received the highest payoff among the nine cells centered on the cell in question. This model is a natural interpretation of the Darwinian model of adaptive success generating reproductive success.

Specifically, each player interacts with her immediate eight neighbors (See Table 5). Without loss of generality, the central player "C" interacts with her eight adjacent neighbors from N1 to N8. She plays the free standing two by two game with each neighbor, but she plays only one strategy per round, either corporate or defect, against all her eight immediate players simultaneously. In short, she is either a pure cooperator or a pure defector and her sphere of interaction is restricted to immediate neighbors. Player C's payoff is determined from the payoff function defined by the stand alone 2 by 2 game of stag hunt, given the plays of her neighbors. Her total payoff is computed by summing payoffs from eight games and the payoffs of her neighbors are determined similarly. In the next round, each agent then surveys her neighbors. For instance, Player "C" observes her own payoff and her eight neighbors' payoffs. She then chooses the action (corporate or defect) that yields the highest payoff among the nine cells. As such, an action with higher performance replaces the initial action, which is selected out of the evolutionary struggle.

Insert Table 5 about here

Trust in this context is measured by the *type* of players, as defined by the 2-period history of her plays [Picker 1997, Epstein and Axtell 1996]. For instance, if the cell player has cooperated in both the previous round and the current round, we code her strategy as 1. All four possible 2-period strategies can be summarized in Table 6:

Insert Table 6 about here

Type 1 player is defined as a 'trust' type player since under this strategy a player chooses to cooperate consecutively. With this set of simulation tools, we set out to investigate whether the players will gradually adopt the 'trust' strategies over time, despite an initially random distribution of strategy types. If the population is increasingly taken over by players with a genetically wired 'trust' strategy, then we are able to show that trust has emerged as a systemic property of the population.

The game stops when all the cells converge to one of the two possible equilibria: [DD] or [CC], which are the two Nash equilibria. We define the first time when the game reaches a uniform action (all "C" or all "D") for all cells as the *converging point*, and call the number of steps/games to reach this point the *converging steps*.

Impact of the initial density of trusting individuals

We quantify 'trust' by defining a trust density index. Density of trust in a particular period is defined as the percentage of trust type players among all the players. We define a trust type player as one who takes a cooperative action in two consecutive periods. In our grid world, there are 6561 (=81*81) players in each period. So in a particular period,

Trust Density Index = Number of players who are "trust" type / 6561 *(1)*

We experimented by initializing the population randomly with different trust density levels. We plot the speed of convergence as a function of different trust densities in Figure 5. In general, the more cooperative the initial population, the faster it converges to an all-trust society (every one in the population belongs to the 'trust' type.) Figure 5 also suggests that when the initial density is below a certain point, it's highly unlikely for the population to converge to an all-trust one (the curve is not tangent to the Y axis.).

Insert Figure 5 about here

While Figure 5 suggests an optimistic outcome for societal evolution of trust, it is intriguing why so many societies (initial distributions of trust behavior) tend to evolve toward cooperation. In particular, societies with little initial trust seem able to overcome this seemingly insurmountable barrier of suspicion.

Insert Table 8 about here

In Table 8, we show the actual density of the first 10 steps for six density levels. These societies often need to go through 'a valley of death' before trusting behaviors take over. Notice that at the first couple of steps, the density of three levels (0.25, 0.15, 0.10) declines quickly. This means that the whole society rapidly degenerates towards distrusting each other. This process is later reversed in an equally dramatic fashion as the remaining trusting souls propagate exponentially by taking over the entire population. This illustrates very well the temporal tradeoff inherent in many natural phenomena – initially the act of defection pays as it exploits the kindness of other trusting players. However, the population is soon quickly depleted of all the trusting souls. This depletion results in a corresponding decline in the number of 'bad elements' before trusting individuals take over again by reaping superior payoffs. However, the process takes almost 30 generations when initial density is at a low 0.1 level.

Impact of initial distribution of trust

From the results of the previous section, we find that the convergence path and speed depend heavily on the initial density of trusting individuals in the population. We further found that the speed of convergence might be completely different between populations with the same initial density. We ran ten rounds with the same initial density of 0.10 and show the number of steps needed before convergence in Table 9. In short, there is a lot of variance in the speed of convergence.

Insert Table 9 about here

This clearly suggests that the speed of convergence not only depends on the initial density, but also depends on how the trust type players are distributed over the space. In this sub-analysis, we further examine the effects of location of these players. We hypothesize that a population with an even distribution of trusting individuals may behave quite differently from segregated populations with the same trust density overall. In order to test this conjecture, we need to develop an index to measure how evenly trust is distributed across the grid.

First we slice the 81*81 grid into 9 blocks (see Table 10), with each block containing 81 players (9 * 9). After initialization, each block contains a certain number of 'trust' type players. If trust is evenly distributed, the nine blocks should contain roughly the same number of trust agents. So any deviation of the number of trust type players from this benchmark indicates an uneven distribution of trust.

Insert Table 10 about here

Therefore, we can measure the evenness of the initial distribution by an index of deviation, which runs from 0 to 9. If a particular initial distribution has N number of trusting players in Block 1, instead of the benchmark M (= total number of trusting type * 1/9, i.e., the expected number in the block), then we add 1 into the index. **/* Huh? Not clear. Explain. */** The same is done for all nine blocks as in (2) below. The higher the deviation, the lesser evenly distribution of trust.

Index Trust Distribution = Deviation (Block1, Block2, … Block9) *(2)*

/* I don't get it. Explain. */

The effect of the initial trust distribution on the evolutionary dynamic is shown in Figure 6, which plots the relationship between evenness of distribution of the trusting types and the speed of convergence, under three different density levels of 0.5, 0.25 and 0.10. The three curves exhibit the same pattern: the higher the deviation, the slower the convergence (more steps to reach convergence). In other words, a more evenly distributed society moves more quickly towards a completely trusting society. Segregation, an extreme case of uneven distribution, impedes the evolution of trust of a society.

Insert Figure 6 about here

It is not difficult to see why distribution matters. A lone cooperator will be exploited by the surrounding defectors and succumb. However, four cooperators in a block can conceivably hold their own, because each interacts with three cooperators; a defector, as an outsider, can reach and exploit at most two. If the bonus for cheating is not too large, clusters of cooperators will grow. Conversely, lone defectors will always do well [not so in our context of stag hunt] **/* What's this? */**, since they will be surrounded by exploitable cooperators. However, by spreading, defectors encounter more of their like and so diminish their own returns.

As in any path dependent process, the actual evolution of such spatial societies is sensitive to initial values. However, the long-term average of the final composition of the population is highly predictable. The only requirement is that each player should not interact with too many neighbors.**/* ???? Everyone has 8. */**

In sum, these results show that trust indeed diffuses under general conditions. The artificial society we simulate here does evolve towards cooperative behavior without external intervention. As such, trust eventually becomes a systemic property of the entire population, despite the fact that initially it only characterizes a small subset of the population. By exhibiting trusting behaviors, both the society as a whole and individuals themselves reap higher payoffs, ensuring that these behaviors will be passed on to the next generation.

Discussion and Conclusion

The subject of trust is fascinating, the focus of much attention, and important both theoretically and for applications, especially in e-commerce. The experiments we report here are admittedly exploratory and much remains to be done. It is remarkable, however, that trust behavior emerges so pervasively in the simulations we undertook. With this in mind, we offer a few comments by way of framing this work and its significance.

Trust is naturally studied in the context of games, or strategic situations, which themselves have been studied from broadly three perspectives. First, classical game theory and its modern descendants have tended to view game theory as a branch of applied mathematics. Games are formalized (strategic form, extensive form, etc.), axiom systems are presented (e.g., utility theory, common knowledge), solution concepts are conceived (equilibria), and results are derived mathematically. The worry about this approach is and has always been that these "highly mathematical analyses have proposed rationality requirements that people and firms are probably not smart enough to satisfy in everyday decisions." [Camerer, 1997] In response, a rich and flourishing field, called behavioral game theory, has arisen "which aims to describe actual behavior, is driven by empirical observation (mostly experiments), and charts a middle course between over-rational equilibrium analyses and under-rational adaptive analyses." [Camerer, 1997]

Our study lies in the third approach (also constituting a flourishing field of research), which Camerer calls "under-rational adaptive analyses" employing "adaptive and evolutionary approaches [that] use very simple models--mostly developed to describe nonhuman animals--in which players may not realize they are playing a game at all." Indeed, since our agents are so lacking in what anyone would call rationality, of what possible relevance is their behavior to the study of trust?

Classical game theory analyses as well as behavioral game theory have shared in common an outlook in which the game players have a psychology of the sort postulated in everyday thought and language. Philosophers call this folk psychology; the artificial intelligence community thinks of it as the psychology of BDI (belief, desire, and intention) agents. Human (intelligent agent) behavior is explained largely by beliefs, desires, and intentions. Behavioral game theory seeks to explain behavior in games by appealing to realistic beliefs, desires, intentions, and computational capabilities of the players. Classical game theory can be understood as investigating behavior under (often unrealistically) ideal cognitive conditions.

In contrast, adaptive or evolutionary game theory has focused on the behavior of simple algorithms and strategies as they play out in strategic contexts. Typically, as is the case for our agents, the players cannot by any stretch of imagination be granted beliefs, intentions, or desires. But they can, and do, play games that are studied from the classical as well as behavioral perspectives. We have found that in the stag hunt game behavior emerges that, *had the players been BDI agents*, would be described as trusting or cooperative. (Analogous results have been found for Iterated Prisoner's Dilemma.) These findings, we think, are interesting and significant in a number of ways.

1) We studied Stag Hunt, a game that has been relatively little studied compared to Prisoner's Dilemma [Cohen et al., 1998]. Stag Hunt represents an alternative but equally interesting context in which to investigate cooperation.
2) It is remarkable that, as in the case of Prisoner's Dilemma, 'cooperative' behavior occurs so robustly. (See [Axelrod, 1984; Grim et al., 1999] for overviews of cooperation in Prisoner's Dilemma.)
3) The dynamics of Stag Hunt (spatial) populations are surprising, particularly the "valley of death" phenomenon.
4) The methodology employed, a form of simulation, is also a form of experimental mathematics. That cooperative behavior emerges in these experiments (in the absence of anything like cognition or rationality) *suggests* that, e.g., the emerging cooperative behavior is simply a reflection of the underlying power arrangements and "correlation of forces" in the strategic situation.
5) The previous remark suggests further that at least some of the behavior of purportedly BDI agents (such as ourselves) might be explained by adaptive or evolutionary processes akin to those that produced the seemingly cooperative behavior in these games. (See [Skyrms, 1996, especially chapter 1] for a similar suggestion.)

Of course, much remains to be investigated. What we have begun here needs broadening and deepening. It is especially important to investigate the space of plausible adaptive mechanisms. In this regard, computational studies of reinforcement learning (broadly construed) have begun to connect with the tradition of classical game theory as it has developed an interest in algorithms for game dynamics (see [Fudenberg and Levine, 1998; Young, 1998, Weibull, 1995]). And then there are possible applications. We close with some brief comments in this regard.

At the individual level, we found that cooperative behaviors can emerge purely as a result of trial and error learning. Trust emerges almost autonomously, without any need for central intervention. This finding has several interesting applications to organizations and businesses in general. For instance, a long-standing discourse in organization theory is the tension between centralization and decentralization. Centralization has the potential to minimize redundancy and waste but runs the risk of over-intervention. Decentralization, however, delegates decision-making power to local entities but in an attempt to optimize selfishly, may be accused of sacrificing global welfare. Our paper provides empirical support to the benefits of decentralization by showing that mutual and seemingly selfish adjustment by subunits does indeed lead to greater global welfare over time. More importantly, this adjustment is achieved without the presence of a centralized authority. This implies that there is no need for complete knowledge of the environment as well.

More specifically, many dynamic interactions in business exhibit qualities that are similar to the game of Stag Hunt. For instance, in joint ventures, two distinct firms have to learn to co-exist and achieve pre-defined goals. While there is certainly common interest in seeing the venture pay off, a conflict of interest also exists when players have incentive to shirk. In a similar vein, in an e-commerce context, the interactions between the 'bricks' unit and the 'clicks' unit also have the distinctive character of a trust game. While they need to cooperate with each other to further the organizational goal, these subunits are also locked in a bitter internal battle as cannibalization becomes inevitable. These diverse business decisions, in particular, simply lack the egoistic character of the

Prisoners’ Dilemma. Instead, they share important common features with a game of stag hunt. As such, they may prove potentially fruitful applications of our research.

At a population level, we find that while trust eventually spreads to the entire population via selection, it does so with an interesting temporal pattern: the valley of death. This finding may have important normative implications from a central planner’s perspective. In particular, it is important for the planner not to be ‘deceived’ by the initial dying down of cooperation. In a sense, this initial drop in trust is a necessary price to be paid for the eventual taking over of the population. As such, while intervention in any form is not needed in general, it should be especially avoided during this valley of death.

Furthermore, our finding at the population level informs the evolutionary theory of the firm as well. While the biological metaphor has been widely used in management, empirical evidence has been sparse. Here we provide a very simple model of biological evolution, where traits – a tendency to cooperate – are passed down genetically. We show that such a simple biological mechanism is powerful enough to change the character of the population, with just a slightly better evolutionary reward. In the context of firm evolution where imitation serves the role of genetic inheritance, this finding translates into a statement about the potent role played by copying the successful template of others. Through simple mechanisms such as imitation, we may begin to explain the homogeneity of observed organizational forms.

To conclude, in this paper, we examine the evolutionary origins of trust by showing how it can evolve at both the individual and the population levels. Although highly simplified and admittedly stylized, our simulation produces some very general and interesting observations. Among possible future work, we may investigate how the dynamics portrayed in our model may unfold in real contexts.

References

Adler, P and S.W. Kwon (2000) "Social capital: prospects for a new concept" Working Paper.

Axelrod, Robert (1984). The Evolution of Cooperation, New York, Basic Books.

Banerjiee, Bikramjit, Rajatish Mukherjee, and Sandip Sen (2000). "Learning Mutual Trust", *in Working Notes of AGENTS-00 Workshop on Deception, Fraud and Trust in Agent Societies*, Barcelona, Spain, pp. 9-14.

Banerjiee, Bikramjit, Sandip Debnath, Sandip Sen (2000). "Combining Multiple Perspectives", *in Proceedings of the International Conference on Machine Learning 2000*, June 29 - July 2 , Stanford University, CA, pp. 33-40.

Camerer, Colin F. "Progress in Behavioral Game Theory" in *Journal of Economic Perspectives*, volume 11, number 4, fall 1997, pp. 167--188.

Camerer C. and T. Ho. "Experience-weighted attraction learning in normal form games". Econometrica, 67. 1999.

Cohen, Michael D., Rick L. Riolo, Robert Axelrod (1998). "The Emergence of Social Organization in the Prisoner's Dilemma: How Context-Preservation and Other Factors Promote Cooperation", Working Paper, University of Michigan.

Epstein, R and R. Axtell (1996) "*Growing Artificial Societies*", MIT Press.

Erev, Ido, Alvin E. Roth (1998). "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria", *The American Economic Review*, Vol. 88, N. 4, 848-881.

Fudenberg, Drew and David K. Levine (1998). *The Theory of Learning in Games*, Cambridge, MA, MIT Press.

Greenspan, A. (1999). Harvard University Commencement Address.

Grim, P., Gary Mar, and Paul St. Denis (1999). *The Philosophical computer: exploratory essays in philosophical computer modeling*.Cambridge, Mass.: MIT Press.

Hannan M. and J. Freeman (1977). "The population ecology of organization", *American Journal of Sociology*,82, 929-964.

Holland, John H. (1995). "Can There be a Unified Theory of Complex Systems?" *in* Morowitz, M and J. Singer (eds.), *The Mind, the Brain, and CAS*, SFI Studies in the Sciences of Complexity, Vol. XXII, Addison-Wesley.

Hu, Junling, Michael P. Wellman. "Multiagent Reinforcement Learning: Theoretical Framework and an Algorithm", *in Proceedings of the Fifteenth National Conference on Artificial Intelligence*, Menlo Park, CA, pp. 746-752.

Kandori, Michihiro, George J. Mailath, Rafael Rob (1993). "Learning, Mutation, and Long Run Equilibriain Games", *Econometrica*. Volume 61, Issue 1, 29-56.

Lave, C. and J. March. (1975). *An introduction to models in the social science*. Harper & Row.

Littman, M. (1994) "Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the Eleventh International Conference on Machine Learning*, pages 157--163, San Francisco, CA, 1994. Morgan Kaufmann.

Macy, Michael W. (1996). "Natural Selection and Social Learning in Prisoner's Dilemma: Co-Adaptation with Genetic Algorithms and Artificial Neural Networks", *Sociological Methods and Research*, Vol. 25, 103-137.

Macy, Michael W. (1997). "Identity, Interest and Emergent Rationality: An Evolutionary Synthesis", *Rationality and Society*, vol. 9, 1997, 427-448.

Macy, M and J. Skvoretz (1996) "The evolution of trust and cooperation among strangers: a computational model" *American Sociological Review*, October, 1998.

McEvily, Bill, Akbar Zaheer (1999). "Bridging Ties: A Source of Firm Heterogeneity in Competitive Capabilities", *Strategic Management Journal*, 20, 1133-1156.

McKnight, D. Harrison, Norman L. Chervany (1996)**.** "The Meaning of Trust", Working Paper, Carlson School of Management, University of Minnesota.

Miller, John H. (1993). "The Evolution of Automata in the Repeated Prisoner's Dilemma", Working Paper, Santa Fe Institute, Economics Research Program.

Morrell, V. (1995). "Cowardly lions confound cooperation theory", *Science*, Vol 269. 1216-1217.

Nowak, M and K. Sigmund (1993) "A strategy of win-stay, lose-shift that outperforms tit for tat in the Prisoners' Dilemma Game" *Nature* 364, 56-58.

Picker, R. (1997) "simple games in a complex world: a generative approach to the adoption of norms" *The University of Chicago Law Review*, 64, 1224-1257.

Rousseau, Jean-Jacques (1755). *Discourse on the origin of inequality*. Translated by Franklin Philip, Original Title "Discours sur l'origine et les fondements de l'inégalité parmi les hommes" Oxford University Press, 1994.

Schelling, Thomas (1978) . *Micromotives and Macrobehavior*. Norton.

Sutton, R and A.Barto.(1998) *Reinforcement learning: an introduction*. MIT Press/Bradford Books

Sandholm, Thomas W., Robert H. Crites (1995). "Multiagent Reinforcement Learning in Iterated Prisoner's Dilemma", *Biosystems*, Vol. 37, 147-166.

Sen, Sandip, Mahendra Sekaran, and John Hale (1994). "Learning to coordinate without sharing information", *in Proceedings of the Twelfth National Conference on Artificial Intelligence*, AAAI, Seattle, Washington, August 1th-4th.

Skyrms, Brian (1996). *Evolution of the Social Contract*, Cambridge, UK, Cambridge University Press.

Warnick, J. (2000) "Learning to trust", Working Paper, University of Pittsburgh.

Watkins, C (1989). "*Learning from delayed rewards*". PhD Thesis, University of Cambridge, England.

Weibull, Jorgen W. (1995) *Evolutionary Game Theory*, Cambridge, MA, MIT Press.

Young, H. Peyton (1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton, NJ, Princeton University Press.

Zaheer, A., B. McEvily and V. Perrone (1998) "Does trust matter? Exploring the effects of inter-organizational and interpersonal trust on performance" *Organization Science*, 9, 2, pp 141-159.

Zaheer, Akbar, Bill McEvily and Vincenzo Perrone (1998). "The Strategic Value of Buyer-Supplier Relationships", *International Journal of Purchasing and Materials*, Module 4, August.

	Cooperate	Defect
Cooperate	3,3	0,5
Defect	5,0	1,1*

Table 1: Payoff matrix of a game of Prisoners' Dilemma

	Cooperate	Defect
Cooperate	5,5*	0,3
Defect	3,0	1,1*

Table 2: Payoff matrix of a game of Stag Hunt

State/action pairs	**First Player**	**Cooperate [C]**	**Defect [D]**
Both cooperate [CC]	Achiever	*	*
One cooperate; the other defect [CD]	Benevolent Loser	*	*
One defect; the other cooperate [DC]	Malicious loser	*	*
Both defect [DD]	Loser	*	*

Table 3: Q function with 4 states and 2 actions.

Initial action	**If observe cooperation**	**If observe defection**	**Reactive strategy**
0	0	0	Always defect
0	0	1	Suspicious doormat
0	1	0	Suspicious tit for tat
0	1	1	Suspicious Quaker
1	0	0	Deceptive defector
1	0	1	Gullible doormat
1	1	0	Tit for tat (TFT)
1	1	1	Quaker

Table 4: 8 Fixed Strategies [0 indicates Defect; 1 indicates Cooperation]

	N1	N2	N3	
	N4	**C**	N5	
	N6	N7	N8	

Table 5: A representation of spatial games

Type	**Strategy**
1	Two consecutive rounds of cooperation
2	Two consecutive rounds of defection
3	Switch from cooperate to defect
4	Switch from defect to cooperate

Table 6: types and their corresponding strategies

	Always Defect	Suspicious Doormat	Suspicious Tit for Tat	Suspicious Quaker	Deceptive /suspicious Defector	Gullible Doormat	Tit for Tat	Quaker
CC	0	0.593	1	1	0	0.45	1	1
CD	0	0	0	0	0.006	0	0	0
DC	0	0.083	0	0	0	0.05	0	0
DD	1	0.324	0	0	0.994	0.5	0	0

Table 7: Evolution of trust in 2-player game of stag hunt

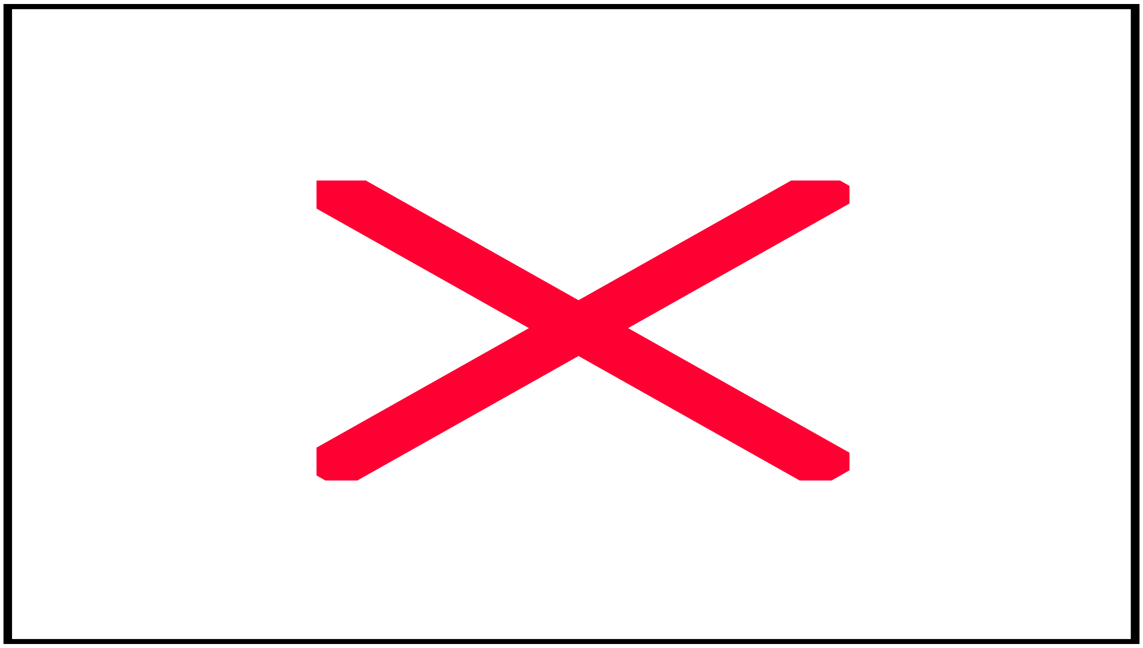

Figure 1: Evolution of Trust for fixed Opponent playing TFT

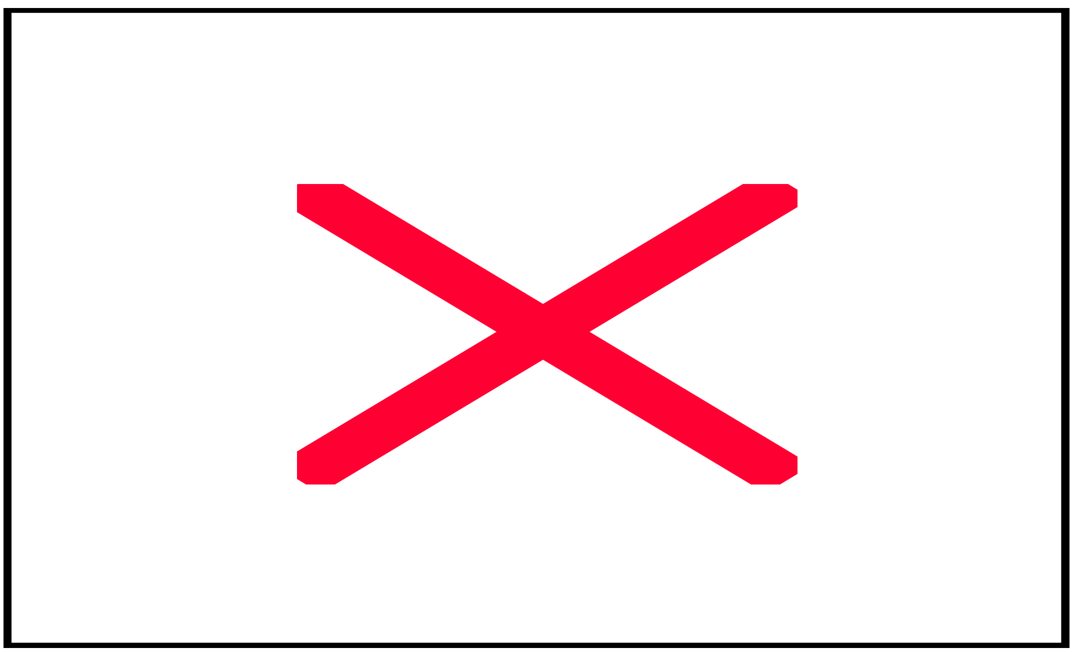

Figure 2: Evolution of Trust for fixed Opponent playing Always Defect

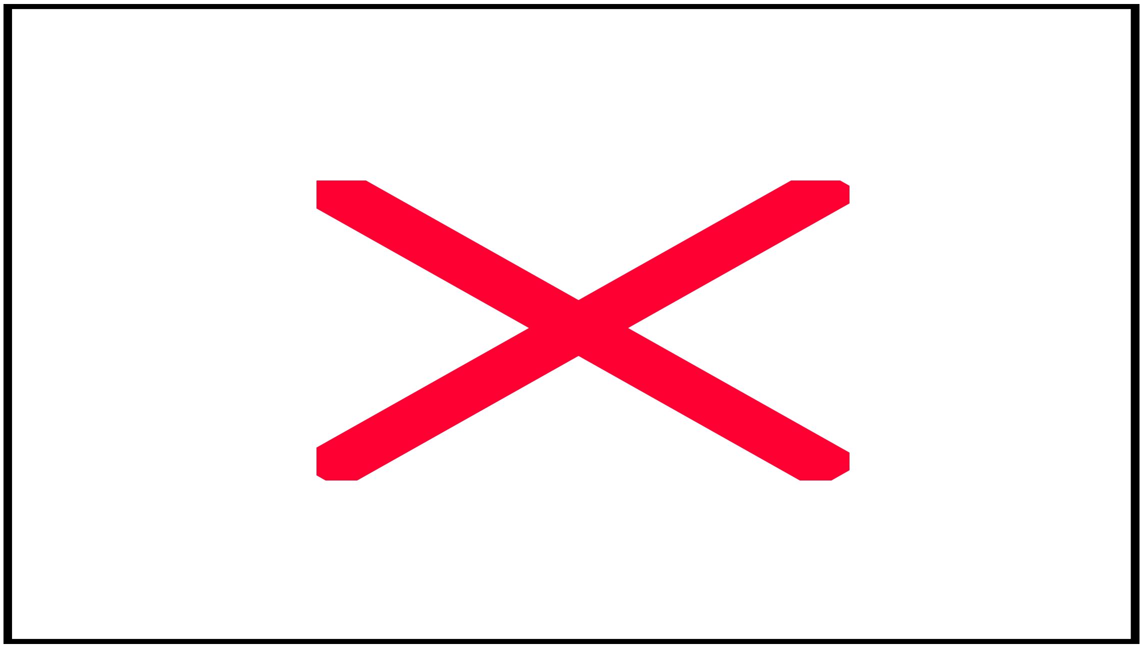

Figure 3: Evolution of Trust for fixed Opponent playing Gullible Doormat

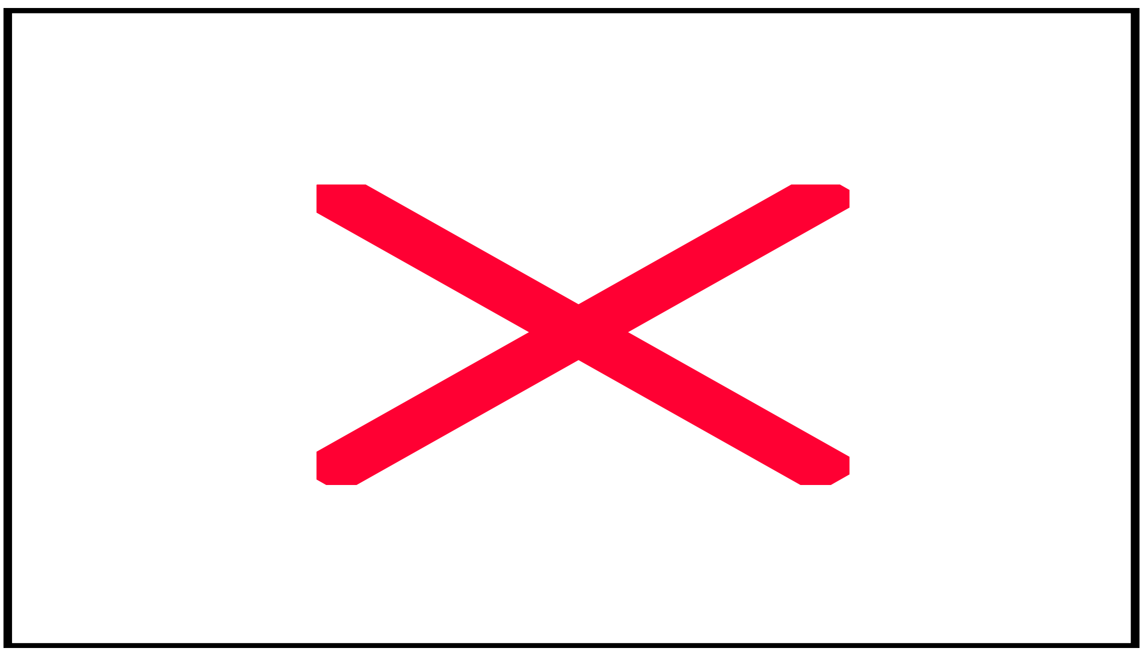

Figure 4: Evolution of trust with two Learners

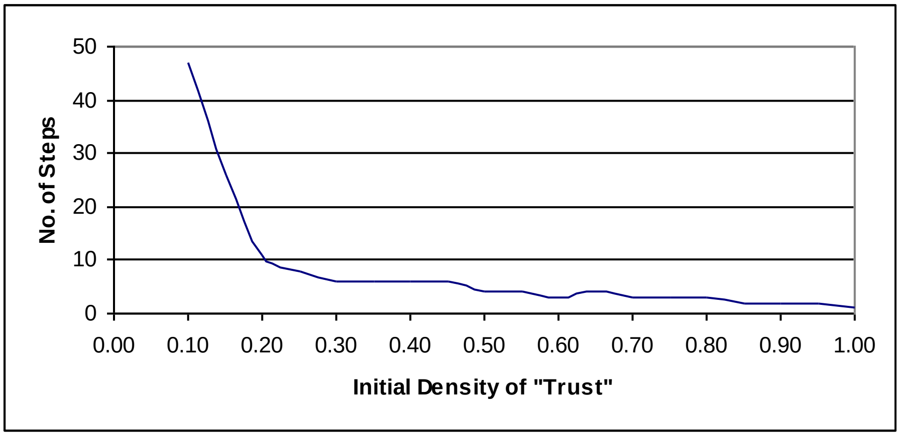

Figure 5: Impact of Density on Speed of Convergence

Steps	Density= 1.00	Density=0.50	Density= 0.25	Density=0.15	Density =0.1	Density =0.05
1	1.00	0.5	0.25	0.15	0.1	0.05
2	1.00	0.59	0.09	0.01	0.01	0.03
3	1.00	0.76	0.26	0.01	0.01	0.02
4	1.00	0.94	0.53	0.03	0.01	0.01
5	1.00	0.99	0.76	0.07	0.01	0
6	1.00	1	0.92	0.14	0.01	0
7	1.00	1	0.98	0.21	0.01	0
8	1.00	1	1	0.3	0.01	0
9	1.00	1	1	0.4	0.01	0
10	1.00	1	1	0.48	0.01	0

Table 8: First Ten Generations of Evolution

Round	1	2	3	4	5	6	7	8	9	10
Steps	26	32	42	37	27	25	39	21	32	13

Table 9: Number of steps to converge in ten rounds with initial density = 0.10

Block 1	Block 2	Block 3
Block 4	Block 5	Block 6
Block 7	Block 8	Block 9

Table 10: Blocks to Measure Evenness in Distribution

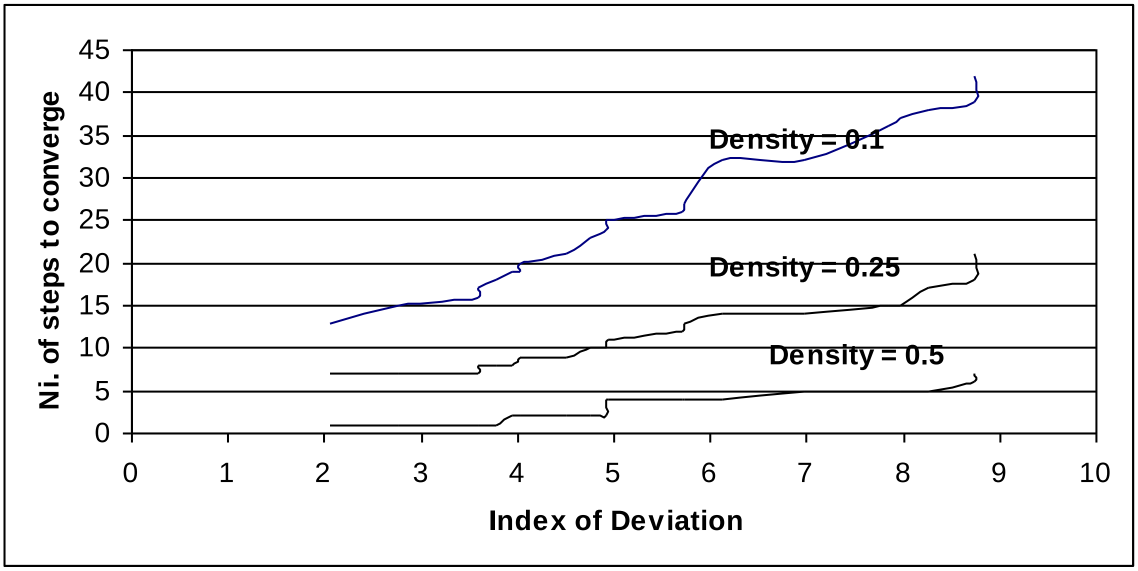

Figure 6: Impact of Distribution on Speed of Convergence

File: staghunt20020323.doc. From: The Jan. 15, 2002 submission.

Date: Mon, 18 Mar 2002 15:49:58 -0500

From: Fang <Fang@management.wharton.upenn.edu>

To: "'Steven 0. Kimbrough'" <sok@grace.wharton.upenn.edu>

Cc: Annapurna Valluri <avalluri@wharton.upenn.edu>,

"Eric Zheng (ericz@grace.wharton.upenn.edu)"

<ericz@grace.wharton.upenn.edu>

Subject: journal website

Here I attach the e-mail from DJ.
"Please make sure your final version
follow the publication guide lines of GIN, at the URL of

http://www.kluweronline.com/issn/0926-2644.

About page 7:

using the following algorithm [Watkins, 1989]:

$$Q(S,A) \leftarrow (1 - \alpha)Q(S,A) + \alpha^* (R + \gamma^* Q(S', A'))$$

where α is the learning rate, γ the discount factor and (S', A') the next period state-action pair.

Come on, the rest of the terms have to be explained!!!!