On Models for Coordination of Activity and its Disruption:
Interim Progress Report, 01/01/03 – 12/31/03

Dr. Joseph Lakey (PI), Department of Mathematical Sciences
Dr. Mike Coombs, Physical Science Laboratory
Kevin Streander, Department of Computer Science
Scott Izu, Department of Mathematical Sciences
Chris Weaver, Physical Science Laboratory

New Mexico State University
Las Cruces, NM 88003

March 2004

1 Recap of Goals for Years 1 and 2

The general project objective is to support the threat evaluation component of Intelligence Preparation of the Battlefield (IPB) for asymmetric opponents by developing a modelling and metrics framework to:

- predict the emergent coordination or disruption of activity in social systems models that
  - employ regulating feedback, often construed as violence, as a means of achieving goals, and
  - use learning to improve performance;
- assess whether the observed systems possess an optimal fitness state that
  - can be attained quickly via a decentralized search strategy, and
  - is robust to small perturbations of rule parameters or environment;
- assess the likelihood that the system can attain such an optimized state.

In Year 1, we undertook to define a conceptual framework for studying emergent coordination in systems in which agents (1) compete for a common, high-valued resource and (2) regulate each others demand through the primitive processes of punishment and retaliation (*see* [16]. These objectives were achieved through the investigation of a simple one-dimensional agent-based simulation. In this simulation, agents competed for a*reporting* resource after completing a sequence of reward earning tasks. A reporting bottleneck had the effect of delaying further reward until reporting was completed. Global system fitness and dynamics were investigated given two parameters governing an agents retaliation against other agents delaying it thorough achieving precedence in the reporting queue. These parameters included agent tolerance of delay, T, and the magnitude of agent retaliation, P. In this model, we were able to identify the components of a predictive analysis, including (1) the system's production capacity, (2) sources of constraint in achieving capacity, and (3) internally driven "stress" arising from inhibiting actions performed by one agent on another with the goal of creating "production" room (defined as P / T). Finally, we identified phase transitions in the stress (P / T) parameter space dividing points where agents frozen in a nonproductive vicious retaliation cycle from points where tolerance and retaliatory restraint permitted productive movement.

In Year 2, we extended the one-dimensional model (Ping I) to two-dimensions (Ping II) and added features to support the learning of risk and reward relationships between an agent's current position (in space and time), configurations of other agents, and their historical behavior at different locations. Through simulation, we have confirmed that, appropriately constrained, the two-dimensional model generates similar behavior to the one-dimensional version. However, the neural network that we have added to implement learning offers the possibility of emulating much more realistic and complex decision behavior. In particular, we can emulate the emotional aspects of decision making that have been found by research in cognitive neuroscience to account for decision phenomena that are not easily accounted for by bounded rationality, such as sacrificial suicide tactics. Given the prevalence of these behaviors in the tactics of asymmetric forces faced with a superpower implementing a strategy of "overwhelming force", it is important for IPB that they be incorporated into threat assessment processes.

2 Simulation Rationale

2.1 Models of human decision making

Our concern in this project is the coordination behavior of asymmetric threats. Decision making by such threats, which may range from insurgents to terrorists to "warlords", may be expected to be abnormal when viewed from a U.S. cultural perspective. Recent experience in Bosnia, Afghanistan and Iraq has demonstrated that asymmetric fighting forces often act in ways that do not appear "rational". Typical deviations that appear paradoxical relative to the U.S. fighting conventions of the clear declaration of "sides" and the optimal use of fighting resources include, for example, (1) sacrificial tactics, including suicide bombings and the continuation of firefights when winning is impossible, (2) dissonant status as an enemy, acting genuinely as a friend in domestic contexts yet as a foe in military contexts, (3) extreme sensitivity to threats, often imagining threats where there are none, and (4) frequent alliance shifting between local ethnic groups or tribal factions, sometimes fighting together against a common cause and sometimes fighting against each other.

These types of behaviors have been hard to incorporate into classical decision models for the Intelligence Preparation of the Battlefield (IPB). The classical approach to threat assessment is to assume that the "enemy" has well defined objectives and a rational strategy to achieving these objectives. By "rational" we usually mean that actions are reasoned to be optimal with regard to some cost/benefit function. Typical functions, for example, might include the benefits of territory gained for the cost of resources expended, including human life.

In order to save the rational strategy assumption when faced with paradoxical behavior, analysts have sought to formulate an array of normative rules and algorithms that define alternative knowledge-based possible worlds where rational cost/benefit analyses in different worlds would explain apparent paradoxes (*e.g.* [7]). However, modern theories in cognitive neuroscience indicate that human decision making involves more than normative rules. Humans understand risk and make decisions in two different ways. One is through the classical analytical "reflective" method, involving explicit, conscious rules and effortful reasoning on the consequences of systems of rules. The other – and biologically more fundamental – method is through automatic "reflexive" processes that are implicit, intuitive and non-conscious ([18]).

From an evolutionary perspective, reflexive decision making is more primitive. It is oriented towards survival and even for humans today is the most natural and fastest way to respond to threat. It is the foundation of Klein's [15] recognition-based decision making paradigm that has become widely recognized as the way military commanders operate, along with others operating in time-critical survival situations. Reflexive decision making is fast but highly dependent on the subject having both experience salient to the current situation and having developed templates with the appropriate access cues. Demasio and others (*e.g.*, [6], [2] [17]) determined that these cues and, thus, the primary human decision making processes inherently emotional (termed "affective"), involving the subject's interpretive associations between situational features, their body state at the time of a decision experience, and the decision outcome.

Reflexive decision behavior is mediated by brain structures known as the limbic system. The limbic system is the part of the brain that surrounds the brain stem below the neocortical layer. A critical structure in the limbic system is the amygdala which has the role of assigning emotional significance to environmental stimuli, whatever their sensory modality. In particular, the amygdala alerts the subject to whether the stimulus is a possible danger, or whether it promises some gain. If it is coded as a danger, it elicits an immediate flight or fight response. Amygdalae of subjects who have been subjected to repeated threat may become highly sensitized to specific threat stimuli, the flight or fight response being triggered by attenuated cues. Indeed, the close coupling between the amygdala and the hippocampus, which mediates memories of alarming stimuli, results in the automatic ascription of meaning to such stimuli. As a result, traumatized individuals are extremely sensitive to cues of possible danger to the point of misinterpretation and over-reaction. With children, repeated exposure to danger leads to permanent sensitization through both chemical and structural changes in the brain (*cf.* Perry [19]). In males, these trauma induced traits are usually expressed as a predisposition to violent behavior (fight) and increased risk taking; in females, the predisposition is towards withdrawal, dissociation and increased caution (flight).

There are two pathways from the senses to the amygdala. The most phylogenetically primitive of these is through a direct link to the thalamus, a sort of sensory relay. Stimulus information is less elaborated through this route, but reaches the amygdala very rapidly, allowing it to initiate a fast response upon which survival may depend. The second pathway routes information from the thalamus to the amygdala through the frontal cerebral cortex. The structures of the frontal cortex are phylogenetically recent and mediate the higher mental processes involved in reasoning and "reflective" decision making. Through this route the amygdala receives much more elaborated information, including the integrated products of the different modes of sensory input. These products may be (1) simple global interpretations of the threat status of perceived entities that balance conflicting threat/no threat cues, or (2) more elaborate interpretations generated through logical inference. Reflective decision making is relatively slow, but gives a more "considered" interpretation of situational data and may also act to adjust or override a reflexive response of the amygdala.

2.2 Toward reactive coordination

According to Ferber [8], any approach to reactive coordination between agents in a signal-based communications system boils down to (1) the use of some form of potential field to determine the movement of agents, and (2) the use of "marks" to coordinate the action of several agents, making it possible to use the environment as a flexible, robust, and simple communication system. However, an agent's coordination behavior may be very different depending upon whether it is operating through its reflexive or reflective system. Some distinguishing characteristics of an individual agents decision making are:

Table 1: Characteristics of decision making

Reflexive	Reflective
Fast	Slow
Primitive emotions (e.g., fear)	Developed emotions (e.g., empathy)
Recognition based	Analysis/logic based
Simple cues	Elaborated cues
Single sensory modality	Integrated sensory modalities
Risk over reaction	Risk under reaction
Simple meanings	Complex meanings

We may, thus, expect that coordination behavior of collections of agents would be different depending upon whether they were working in reflexive or reflective modes, or depending upon distributions of reflexive and reflective agents in the systems. Moreover, we would expect differences to manifest whether the agents were coordinating aggressively, in conflict, or affiliatively, in collaboration. For example, reflexive agents in conflict may be expected to become rapidly locked into retaliation cycles given minimal local provocation; reflexive agents in collaboration may be expected to become rapidly locked into support cycles with minimal local reward. On the other hand, reflective agents in conflict may be expected to slowly converge to optimal, resource preserving retaliation cycles given maximal provocation integrated over time and space; reflective agents in cooperation may be expected to slowly evolve support relationships that determine level of cooperation based on level or reward.

In its most primitive form, the Ping II simulation model discussed here is intended to investigate the possibility of coordination when communication among agents is minimal. Given our previous focus on conflict behavior in Ping I, Ping II agents are aggressive/reflexive. We are still investigating principled design of reflective decision making in such a system. In Ping II, the role of Ferber's "marks" is played by "tags" that record nothing but when a site on the game board was last visited by another agent and used as a resource, and by whom: no other information, such as the direction of movement of the last visiting agent, is associated with the tag. In this sense, the agents react like reflexive agents.

By analogy with energetic models, it makes sense that Ferber's potential field should have an attractive and a repulsive component. The attractive component can be defined in terms of a collectable or controllable limited resource, while the repulsive component has to be modelled in terms of competition or danger attached to greedy behavior. Competitive systems can be inefficient when constraints imposed on agents pursuing individual reward are poorly aligned with conditions needed for good performance at the system level – the familiar "tragedy of the commons". Wolpert and Tumar [26] suggest that, at least in certain dynamical systems, it is possible to transform local utilities in such a way as to align them to global utilities. When agents follow a homogeneous set of rules, agents who are obtaining better than average (rather than capacitary average) reward have no reason to alter their behavior, although agents performing worse than average do. However, as poorly performing agents attempt to increase their reward they can trigger waves of retaliation that can bring down the system, leading to all agents doing worse and thus under-utilization of capacity.

Thus, in order to effect better global performance in conflictual systems without external influence, it makes sense that agents should have recourse to influence each others actions. In turn, agents should have the ability to learn that greedy impulses can have costly consequences. Perry (2001) defines the resource conditions under which this is most likely; the predictable, safe and stable conditions given below. Under these conditions, such globally optimal behavior is likely to occur with both reflexive and reflective decision making. With other resource conditions, optimal behavior will require considerable reflective control making globally optimal behavior less likely. This is particularly the case with populations of individuals traumatized from childhood.

Table 2: Resource conditions for reflexive/reflective decision making

	Resource-surplus	Resource-limited	Resource-poor
Social-Environmental Pressures	Predictable/ stable/safe	Unpredictable/ novel	Inconsistent/ threatening
Prevailing Cognitive Tone	Abstract/ creative	Concrete/ superstitious	Reactive/ regressive
Prevailing Affective Tone	Calm	Anxiety	Terror
Systematic Solutions	Innovative	Simplistic	Reactionary
Focus of Solution	Future	Immediate Future	Present
Rules, Regulations, Laws	Abstract/ conceptual	Superstitious/ intrusive	Restrictive/ punitive

Resource conditions will also affect perception of risk, what will be learned about a situation and what will be perceived as reward. For example, agents adapted to a resource poor environment will be very risk tolerant, will be rewarded by a minimal reduction of threat from other agents, and will only learn aspects of the threat environment temporally associated to the threat event. Agents adapted to resource surplus conditions will be risk avoiding, future oriented and are likely to use long-range reflective reasoning about the optimal response to a threat.

3 Ping II Design Considerations

Ping II is an example of a *mobile social network* in which agents move along the edges of a graph according to some set of update rules. Reward is distributed at the nodes of the graph and is updated according to some allocation rule. Each individual agent seeks to collect reward at an optimal rate. Local update rules and reward allocation should conspire to provide satisfactory system performance. The two goals – local and global optimization – can be in conflict because, by collecting reward as rapidly as possible, an agent might diminish what remains for other agents in its wake. Thus we see an interplay of four key features of a social network at this level of abstraction:

- Topology of the underlying graph on which reward is allocated and collected
- Law(s) by which reward gets allocated
- How agents interact with one another
- Agent decision rules for where to move next

The simulation framework that we present here accounts for some of the recent learning theory of social games, at least in rudimentary ways; it will do more so as the program continues to evolve.

In order to keep topological issues reasonably simple, we work specifically with square (planar) grid models in which agents at interior nodes can move in one of four directions (N,S,E,W) while agents at boundary nodes are prohibited from moving in a direction that would leave the grid. We hope to consider more complicated topology in future simulations but this is not a top priority. One important feature of grid geometry is scalability: basic quantities like *density* (agents per node) are independent of system size. On the other hand, despite the simplicity of grid geometry, boundaries do pose nontrivial constraints on systems - particularly for small grid size in which the ratio of perimeter to area is nontrivial.

Social networks are often phrased in game-theoretic terms in which payoffs are determined by the collective actions taken by agents during play. Rational agents then react to one another's actions and strategize over what to do next. Thus agent actions have an immediate consequence: the collection of actions determines each agents payoff. But agent action also affects the evolution of the system because agents respond to one another's actions. In mobile social networks individual agents seek to optimize reward over time. Thus a successful agent will either be lucky, or will be endowed with sufficient rational powers to formulate a plan of action consistent with expected actions of other agents.

Besides the notion of individual reward, however, in a social network there is also the notion of *common good*. In social games, the instrument for the achievement of common good is *cooperation*, though the particular meaning of this term depends on context. By *coordination*, on the other hand, one typically means the use of interaction among individuals to carry out collective activity. This can mean to carry out a specific collective action that could not be carried out by a single individual. But in a reward based system it can also mean, more generically, the use of interaction to optimize collective reward.

In social games, optimality is phrased in terms of Nash equilibria: each agent is optimizing its actions (in expectation) given the (expected) actions taken by others. Technically, *cooperation* usually just refers to one option of play. A game is then characterized as a *cooperative game* if a global Nash equilibrium occurs when most or all agents "cooperate". This is often not the case. For example, in minority rule games, each player casts a "vote" (say by choosing 0 or 1). Only those agents voting in the minority – that is, not "cooperating" – are then rewarded. This particular game is "noncooperative" in the sense that global payoff is then maximized when there are as few cooperators as possible.

The minority game is "ill-constrained": there is no collective strategy that benefits every player at once. Nevertheless, under suitable learning rules an agent can learn to optimize its own *average payoff*. Global payoff is also optimized, on average, when all agents learn to do so. As gents to not actually *interact*, outcomes depend only on laws for allocation of reward and how agents decide which action to choose.

In mobile social networks reward is spatially distributed. While decisions on which action to take can be cast in game theoretic terms, in comparison with social games, *rationality* now has an even more complex meaning because decisions depend on a context – not just on 'what agents tend to do' as in social games, but also 'where the other agents are'.

Systems in which reward is distributed over some graph structure can be formalized in the language of 'landscapes' and we will return to this below. For now, observe that *learning* must take configuration into account. Nevertheless, there is an important lesson from recent developments in social game theory that awaits analogous formulation in the context of mobile social networks, namely that, in the absence of complete knowledge of the decision rules of other agents, a rational agent can benefit from *probabilistic behavior* based on expectation of the behavior of other agents. Moreover, this *hypothesis based decision making* relies, on optimality in probability, on a principle of *non-exclusion*, namely, that no hypothesis should be entirely excluded [10]. This principle has not yet been embodied in Ping II because of the need to understand first the role of the various system parameters in reproducible simulations.

Interaction is the hallmark of a social system. As Ferber [8] suggested, *tags* play a fundamental role in communication networks. The simplest form of a tag at a site simply keeps track of when the site was last visited and or who last visited the site. Further state information might also be left with the tag.

Perhaps the main aspect that differentiates Ping II from other mobile social networks is that an agent can take an action to prohibit another from collecting reward, under certain conditions. Coupled with the law for allocation of reward, when rational agents are able to form appropriate hypotheses regarding the consequences of their actions in light of this form of coercion, it is possible that a Ping II network can exhibit a distribution of reward collection that is consistent with a reasonable notion of coordination – namely that agents (i) avoid getting in each others way – insofar as this is possible and (ii) by doing so, global achievement of reward can – in principle – be optimized when individual reward is distributed in a homogeneous way.

The remainder of this report is organized as follows. First, in Section 4 we provide a mechanical description of the Ping II simulation. Next, in Section 5 we report mathematical observations and results relevant to Ping II and related systems. In particular, we identify an important aspect of the structure of allocation of reward in steering system behavior. This is followed in Section 6 by further simulation results and observations that point to additional modifications that should be studied and theoretical directions that need to be pursued (Section 9). In order to describe these directions further, we summarize recent literature in two specific areas: the first is learning in the context of social games and mobile agent networks; the second is landscape theory, in particular, combinatorial and spectral theory in the context of landscapes on graphs.

4 Outline of Ping II Mechanics

4.0.1 Agent logic

In Ping II, a system of N agents inhabits a two-dimensional square grid of size $K \times K$. Like Ping I – the one-dimensional Ping model discussed in [16] – Ping II is a *threshold-delay* model. The system is provided with a 'random' seed that determines initial values of all agent and system parameters. Agents are assigned an identification number $i \in \{1, \ldots, N\}$. We refer to the i-th agent as 'agent A_i'.

In each *time step* the system of agents undergoes a two-stage process. Each stage, in turn, can be broken down into several subroutines. Agents take their turns to act according to a (randomly initialized) queue. Those agents that are most delayed go to the back of the queue. Agents that get to move first in a time step have first opportunity to gather reward that is within reach. For the purpose of this report, agents are *homogeneous* in that preset parameters that affect agent decisions are the same for all agents.

In the first stage of each time step, agents move (if they can) and, if not delayed, grab reward and update *tag value frustration*. At this point the neural risk calculator (see Section 4.3) is also updated. In the second stage, agents check whether they are frustrated and, if not delayed, they ping and update all *ping frustration* values. This process per time step is broken into two stages in order to avoid conflicting assignments of parameter values. On the other hand, this two stage method also introduces a *reaction lag* (ping frustration is not *felt* until one time step passes).

Here is a more detailed description of A_i's logic.Further details of particular quantities will be given below. In stage 1:

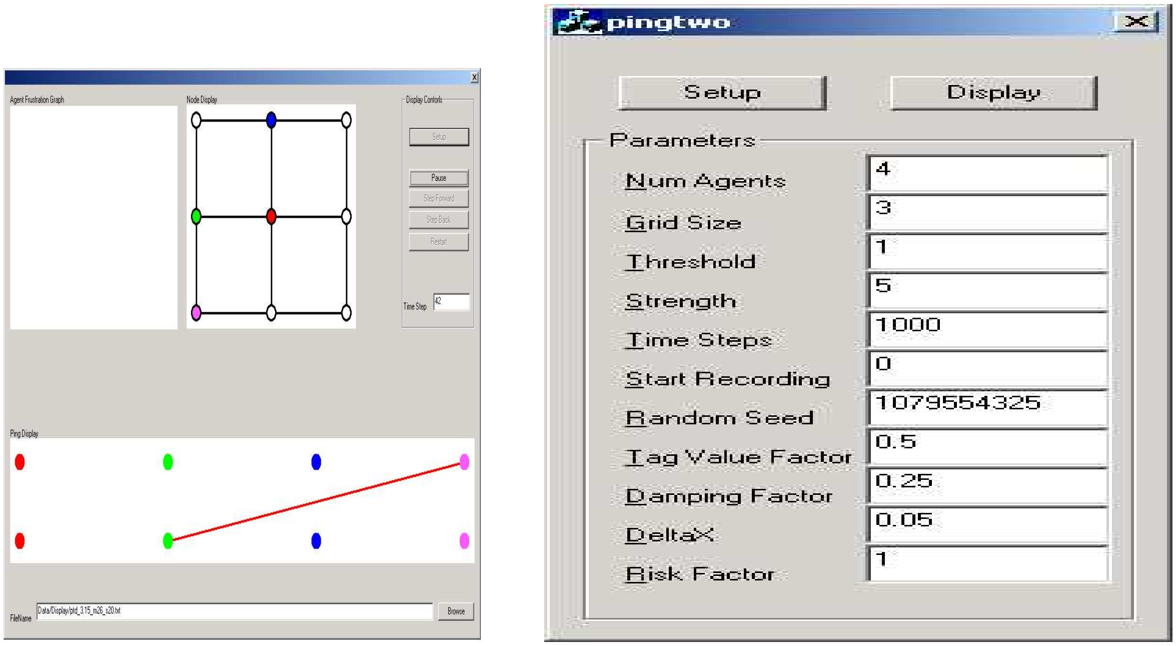


Figure 1: Ping II Display window

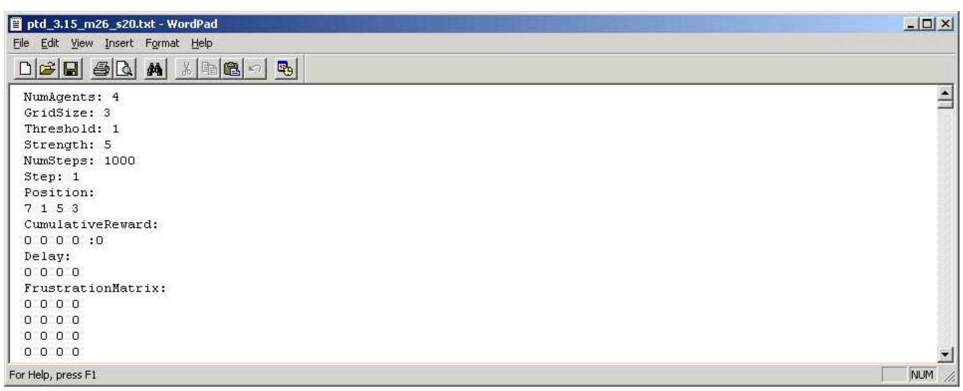


Figure 2: Ping II research output file

- Check if there are neighboring sites open.
- If a site is open then:
  - For each open neighbor, compute (Reward - Normalized Risk).
  - Move to open site that maximizes this quantity.
  - Globally update A_i's location.
- If A_i believes that it owns its current node then:
  - Check owner of current tag.
  - If owner is A_j ($j \neq i$) then increment f_{ij} – A_i's frustration counter for A_j.
- If delay counter is greater than zero then decrement delay counter. If delay is zero then:
  - Collect reward at site (one minus current tag value) and update cumulative reward.
  - Reset tag value to 1/tvfact and (if necessary) tag ownership to A_i.

After every agent has taken its turn in stage one, every tag value is updated by multiplying by tvfact. Here tvfact is the Tag Value Factor item in the setup menu in Figure 1.

Now we can move to Stage Two – the so-called *ping stage*. Again, agent logic is carried out in the order of agents in the queue.

- If A_i is frustrated (sum of frustrations exceeds threshold) and not delayed then:
  - Check for neighbors.
  - Distribute pings among neighbors.
  - Zero out frustration vector.
  - Update risk neural net (see below).
- increment delay by total number of pings received and increment f_{ij} for each ping received from A_j.

Several of the items introduced in this description of the agent logic merit further clarification.

4.0.2 Neighborhood

This consists of those nodes on the grid that are unit distance from A_i. Here, distance is path distance, with each edge counting one unit. On a bounded grid, each interior node has 4 nodes in its *von Neumann neighborhood*. Each boundary node that is not a corner has 3 neighbors and each corner has two neighbors. The dynamics depend heavily on this neighborhood structure.

4.0.3 Strength and pings

An agent's pre-set strength S is the number of ping delays the agent distributes when it pings. Pings are distributed by A_i as evenly as possible among those agents in A_i's neighborhood. If the number of agents in A_i's neighborhood does not divide S then A_i pings first any neighbor A_j that frustrated A_i the most, and randomly in the case of ties.

4.0.4 Delay, frustration and threshold

Agent A_i keeps track of delay by means of a counter D_i. Any time A_i is pinged, D_i is incremented by the number of pings received. If $D_i > 0$ then A_i is said to be in *delay mode* or simply *delayed* or *passive*. Otherwise, $D_i = 0$ and A_i is said to be in *active mode* or simply *active*. An agent that is delayed is not allowed to ping or collect reward, but its other processes are unaffected.

Agent A_i keeps track of frustration with a vector F_i of weights f_{ij} – frustration caused by A_j. There are two sources of frustration. First, when A_i is pinged by A_j, A_i increments f_{ij} by the number of pings received from A_j. Secondly, when A_i arrives at a site last inhabited by A_j, A_i increments f_{ij} by one unit.

In the ping stage, if $D_i = 0$ then A_i checks whether its accumulated frustration $\sum_j f_{ij}$ exceeds a pre-set threshold T. If this is the case, then A_i pings – as above – if possible (A_i's neighbor set must be nonempty). Upon pinging, the frustration vector F_i is reset to zero. If A_i has no neighbors then F_i stays fixed.

It is worth pointing out that frustration can be unbounded. This happens quite often when the grid is uncongested. An agent does not actively seek out others to ping. To the contrary, it tends to avoid contact with frustrated agents (seek Section 4.3). As agents only zero their frustration upon pinging, and since pinging requires having neighbors to ping, agent frustration can increase unboundedly in uncongested grids where avoidance is possible. Whether frustration is unbounded, however, is immaterial. All that really matters is whether $\sum_j f_{ij} > T$. The relative magnitude of the f_{ij} also matters in the case of distributing 'leftover' pings.

4.0.5 Tag value and reward

When agent A_i *in active mode* visits site s, A_i leaves a marker or *tag* at the site. Tag values are updated globally at the end of stage one. For those sites that are occupied, the identity (index) of the occupying active agent is recorded. For those sites that are not actively occupied at the end of stage one, the identity of the last active agent that occupied the site is recorded. If A_i is active and forced to remain stationary due to lack of open neighbors, then the tag value remains equal to one. If a site open at the beginning of the stage remains open then its tag value is reduced by the Tag Value Factor *tvfact*. The tag value thus automatically determines the last time of active occupation.

Agents seek to maximize their cumulative reward. For Ping II, reward is allocated at each site on the grid according to the following rule: The reward at site s is equal to $1 - (\text{tvfact})^t$ in which tvfact is the universal tag value factor and t is the number of time steps since the site was last actively occupied.

In order for a site to have available reward at stage one of a time step, the site must not have been actively occupied at the end of the previous time step. A consequence is that if the number N of agents is more than half the number K^2 of nodes, then at least one agent will be unable to collect reward in a given time step. We will refer to such a grid as being *congested*. Clearly, the more agents the more congested. Of course, congestion can and often does arise *locally* on the grid even if $N < K^2/2$.

This rule for updating reward can have interesting consequences. In particular, in a congested board, pinging puts agents in delay mode. Since delayed agents cannot update tag ownership, this enables reward to accumulate at passively occupied sites at which it would have otherwise been unable to accumulate. Consequently, in congested grids overall system performance can actually benefit from ping induced delays.

4.0.6 Where to go next

Agents actively seek reward. To optimize reward collected over time, agents need to make good decisions about where to move next. There are many ways to model such decision processes, one of the more popular ones being reinforcement learning such as Q-learning. Such methods work quite well in the long run but are not always accurate models when social interaction comes into play.

The decision process in Ping II is predicated on the hypothesis that, in many social environments, reward brings out more impulsive behavior than does risk avoidance, though the two sides must be balanced in order to make an appropriate decision among choices of actions. In Ping II, a neural network architecture associates a *risk* to moving to a given site. The risk is a number between zero and one computed as described in Section 4.3. The agent decision process is then described as follows:

- If a neighboring site s is open, compute: $V(s) = \text{reward}(s) - \alpha * \text{risk}(s)$
- Move to open neighbor that maximizes $V(s)$.

Here α is a *risk multiplier* which balances an agent's greed against its reticence. Reward is defined in terms of tag value as above. When $\alpha = 0$ an agent does not take risk into account. The effect of varying this parameter in grids possessing different levels of congestion is currently under investigation. Preliminary findings indicate that, for good system performance, it is reasonable to choose a small value of α when the system is not congested and a large value when the grid is densely populated by agents.

In principle, the quantity $\max\{V(s) : s \text{ an open neighbor}\}$ could be negative, indicating that, under current conditions, it is safer to stay put. Even so, the agent still must move if possible. This rule forces agents to *explore*, even in a risky environment, to increase the system potential for finding better patterns over which to collect reward.

4.0.7 Update order

Because scarce resource (reward) is an integral part of Ping II, it is impossible to have agents carry out their processes simultaneously in parallel. This is significant when comparing Ping II to certain N-player games in which simultaneous action presents a primary obstacle to learning. We could impose a similar dilemma here, for example, by holding an auction for available sites, but this would only add complexity to an already quite complicated system.

Agents thus take their turn in order of an update queue. This queue is determined at the end of the Ping step. Those agents that are delayed the most go to the end of the queue. Ties are broken by agent identity, namely, if $D_i = D_j$ and $i < j$ then A_i precedes A_j in the queue. This tie-breaking tends to give a slight advantage to A_i; however, in any particular initial condition (but not on average) initial settings tend to outweigh the effect of this advantage.

The order in which agents carry out their processes can also be a factor in system performance. One noteworthy point is that ping frustration is not computed until the end of the ping step. This means, essentially, that there is a lag of one step between when an agent is pinged and when the ping registers as delay or frustration. In uncongested networks this can help to enable agents to avoid one another.

4.1 Dependence on initial conditions

Except for initial random seeds, ping is essentially deterministic (except for minor tie-breaking events). This implies that limit cycles can form – indeed, they are encouraged by the risk calculator – particularly in uncongested networks. The structure of these cycles, however, can depend sensitively on initial conditions. Because of this, it is important to average over a number of initial random seeds in order to justify qualitative statements regarding the behavior of the system over ranges of parameter values.

4.2 Boundary effects

Behavior of social networks is also notoriously dependent on interaction, or graph structure. The differences between such structures can be seen in the following simple example of two 3×3 grids – one 'toral', one 'bounded', populated by three agents in Figure 3. In the toral grid, the three agents wrap around on the third step. When the agents move in parallel lanes (and hence avoid one another so that there is no pinging) then, the reward collected by each agent in each time step in the toral case is $1 - (\text{tvfact})^3$. In the corresponding bounded grid case agents must turn around at the boundary. In this case the average reward collected over a 4-step round trip is $1 - .5[(\text{tvfact}) + (\text{tvfact})^3]$ which smaller than the average toral rate. Both behaviors are optimal. Generally speaking, the effect of the boundary is diminished with grid size.

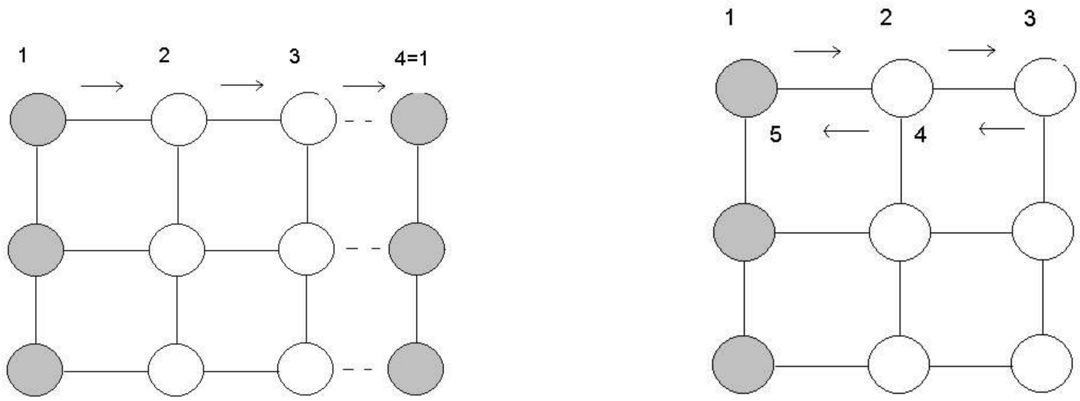


Figure 3: Toral grid versus bound grid

4.3 Role of neural architecture: centralized vs shared info

In what follows we outline a neural architecture for associating a value of risk to moving to a site. The calculator is a neural net with hidden layers. It should be emphasized that the calculator is *first generation.* Our immediate goal is to determine when and to what extent risk aversion – even in a naive form – has beneficial long term effects.

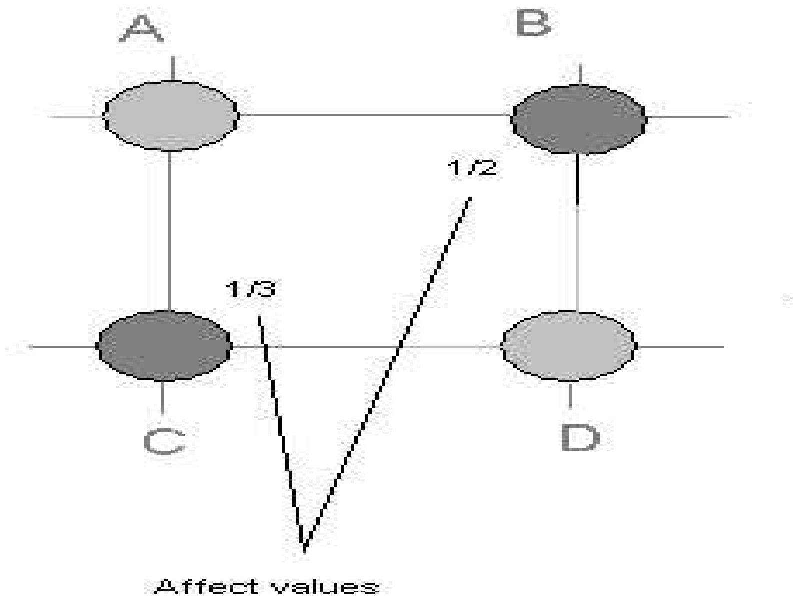
In Figure 5 the nodes A, B, C, D are indexed by the actual grid sites for the simple 2×2 grid in Figure 4. The input neurons at the top are indexed by the neighbors of those sites. Each top node actually references a vector - a numerical entry for each agent that could be on the node. At a given time step, the net checks which agent occupies the node and multiplies the corresponding weight by a value called the *agent affect value* that roughly measures some amount of danger associated to the agent being on that node. The value is determined by:

$$\text{Affect}(F, D, T) = \begin{cases} \exp\big(-D(1 - \frac{F}{T})\big), \ (F < T); \\ \exp(-D) \ (F \geq T) \end{cases}$$

Here F denotes the cumulative frustration experienced by the agent on the given node, while T denotes that agents threshold and D its delay. The risk calculator thus uses full knowledge of these values, as opposed to some sort of guess. The affect of an agent at a node lies between zero and one. It is small if the agent occupying the node is largely delayed or relatively unperturbed. The value input to a node in the second row is the sum of the weights times affects of neighbors. It gives a rough idea of which nodes are dangerous to move to.

Values at nodes in layers (rows) 3 and 4 are computed by summing values of nodes in the preceding row multiplied by corresponding weights in $[-1, 1]$ (to model reinforcement and inhibition) indicated by connecting edges. The resulting value σ is then passed through the sigmoidal function $\frac{1}{2} + \frac{1}{2}\tanh(\sigma)$ to yield a node value in $[0, 1]$. Roughly speaking, row 3 encodes geometric information about the dangerous neighborhoods. Row 4 accounts for all of the influences that nodes have historically had on each other. As is typical practice in neural nets, the extra nodes in layer 4 help to prevent chaotic oscillation – an obstacle to convergence. The final output is a global risk value denoting the risk attached to moving to the node in question. When it comes an agents turn to move, the agent computes the risk associated to each of its possible moves, and chooses to move to a site that maximizes $\text{Tag} - \alpha\,\text{Risk}$.

The weights along each edge are updated whenever an agent pings on a given node. In this case, each weight is updated according to the rule $w_{\text{old}} = \tanh(x) \mapsto w_{\text{new}} = \tanh(x + \Delta x)$. Here Δx is the menu



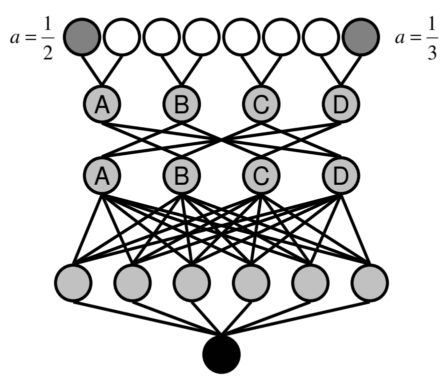


Figure 5: Ping II risk neural architecture for grid in Figure 4

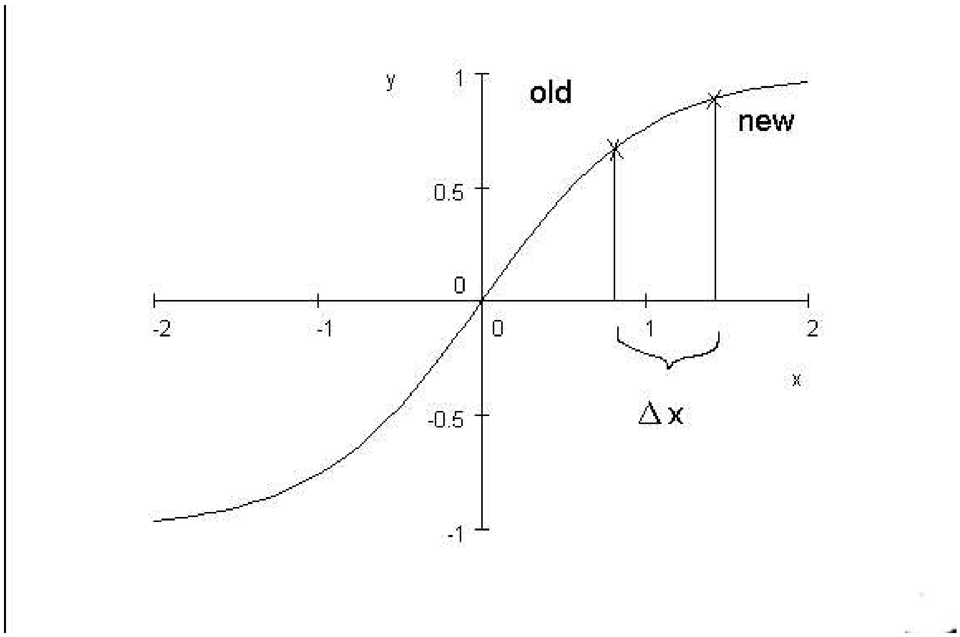


Figure 6: Update rule for weights

parameter in Figure 1. Again - one could allow more sophisticated update rules in which weights at different levels are updated at different rates. Perhaps more significantly, the weights in the risk calculator are monotonically increasing. That is, the risk calculator is increasingly sensitized with every ping. A different possibility would be to *habituate* the system – decrementing the risk by some forgetting factor each time there is no pinging. In view of our primary goal of assessing the impact of naive risk aversion, and due to complexity issues, we have omitted these more subtle factors from the initial design. They will be included in future versions when we have sorted out optimal learning strategies for agents.

In the absence of a habituation factor, the weights will grow monotonically. Thus, if a weight is randomly initialized to a large value, an agents will never learn that its role in evaluating risk might be unrealistically pessimistic. There are two ways in which to alleviate this pessimism. The first is habituation, as noted. The second is to use a *risk factor* α. Simulation data illustrates that in the uncongested regime ($N << K$), a much smaller factor α is warranted. That is, uncongested grids are relatively non-risky and agents need not be overly cautious. Conversely, in congested grids, avoidance is more difficult: Agent frustration is likely to build more quickly and pinging is more likely to ensue as a result.

5 Theoretical Observations

5.1 The role of reward

As noted above, Wolpert and Lawson [25] considered the problem of how to "initialize/update the payoff utility functions of the individual processes so that the ensuing behavior of the entire collective achieves large values of the provided "world utility" as one of the major concerns in distributed multi-agent systems. Most work on payoff assignment is phenomenological. Our goal here is to develop some basic principles regarding how reward ought to be defined so as to minimize conflicts between local and global performance.

In Ping II, optimizing collective reward is consistent with optimizing individual reward *on average*. In fact, because the agents behave homogeneously, it makes sense that individual payoff should favor average behavior. Of course, in real systems, agents will not be perfectly homogeneous. Nevertheless, it is important to understand the homogeneous case first. In Ping II, the reward function located at a site is a monotone function of time and it is stationary in the sense that reward grows at the same rate at each site. Saying that the function favors average behavior is the same as saying that the value of a function at the average of several times exceeds the average of the values at the particular times. That is, *reward should be convex.*

In Ping II, the reward available at a site obeys the rule $r(t) = 1-\beta^t$ in which t is the number of time steps since reward was last collected at the site. Previously, we regarded β as the tag value factor tvfact $\in (0,1)$. In this way tvfact^t is construed as a *decay of influence*. However, $\beta = \text{tvfact}$ can also be thought of as a relative rate of replenishment of a resource.

5.1.1 Reward as replenishment

Suppose that reward $r(t)$ at a site t clicks after the last collection be replenished at a (logistic) rate proportional to $1 - r(t)$, with proportionality constant $\gamma \in (0,1)$ called the *replenishment rate.* Then

$$r(t+1) = r(t) + \gamma(1 - r(t)). \tag{1}$$

When $0 < \gamma < 1$ and $r(0) = 0$ one has, upon iterating (1):

$$r(t) = \gamma \sum_{j=0}^{t-1} (1-\gamma)^j \ = \ \gamma \frac{1-(1-\gamma)^t}{1-(1-\gamma))} = 1-(1-\gamma)^t.$$

Setting $\beta = 1 - \gamma$ then one has

$$r(t) = 1 - \beta^t \tag{2}$$

which is the expression we obtained from viewing reward in terms of tag value decay.

5.1.2 Convexity

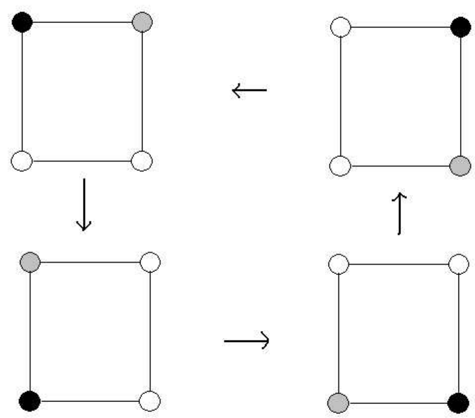
In this sense $r(t)$ behaves much like the saturation phase of a logistic growth model. What is important, mathematically, about this function $r(t)$ is its *midpoint convexity*, namely, $\frac{1}{2}(r(t-1) + r(t+1)) = 1 - \frac{1}{2}(\beta^{t-1} + \beta^{t+1}) \leq 1 - \beta^t = r(t)$, which follows from $1 < \frac{1}{2}(\frac{1}{\beta} + \beta)$ for $0 < \beta < 1$ since $\beta + 1/\beta$ is decreasing on $(0,1)$. This midpoint convexity gives us a best bound for the rate of collection of reward by a system of N agents populating a graph of K nodes.

Theorem 1 *Let G be a connected graph that can be decomposed as a disjoint union of N cycles C_ν, $\nu = 1,\ldots,N$. Suppose that reward is allocated at each node with the fixed convex function $r(t) = 1 - \beta^t$ $(0 < \beta < 1)$. Then the rate of reward collected by N agents, one per cycle, is maximized, among all such possible decompositions into N disjoint cycles, when the variation of the lengths of the cycles about their average E is minimized.*

By a *cycle* here we mean, importantly, a simple closed path. That is, the path has no self-intersections or sub-cycles. In a sense the statement of this theorem is harder than its proof. Nevertheless, it lends some insight into the role of the reward function. Some further comments are worth considering before we proceed to the simple proof. The first is that, though we have phrased the result for logistic replenishment, the theorem really applies to any convex reward function.

The question of when a general graph G is decomposable into disjoint cycles is an interesting problem in its own right. There are further important considerations for Ping II – even in the case of a bounded grid – that we will investigate further. The theorem says that local and global reward constraints can be perfectly aligned when the number of agents N divides the grid size K^2. Then the ideal cycle length per agent is K^2/N. In the case of 3 agents on a 3×3 grid, there can be simple topological obstructions to the desired decomposition into disjoint cycles. In such a case there are two alternatives that can lead to optimized

Figure 7: One agent following another



behavior: either agents move along disjoint (noncyclic) sets of nodes or else agents move along overlapping nodes but temporally avoiding one another's neighborhoods (and thus avoiding pings). In both cases, it is difficult to formulate optimal movement strategies. In fact, the simple example of two agents on a 2×2 grid illustrates that overlapping cycles also allow for the possibility of optimal collection of reward. Figure 7 illustrates suboptimal behavior. (Actually, this sequence of movements will not realistically be realized: the grey agent will not follow the black agent directly since grey then will not collect any reward: this pattern could only result from unrealistic fear of retribution.) On the other hand, Figures 8 and 9 illustrate optimal behavior.

The previous remarks refer only to uncongested graphs in which either agents own disjoint sets of nodes or at least always possess empty sets of neighbors. The theorem does not address the case of *congested* networks. In that case pinging can actually serve to improve system performance by decreasing the number of active agents, and we are continuing to investigate the role of phase relations in pings there in terms of optimizing global performance.

Proof of Theorem 1. One only needs to show that if a pair of agents A_i and A_j have cycles C_i, C_j with lengths $\ell(C_i) < E < \ell(C_j)$ then the global payoff would be increased if one could replace C_i and C_j by a new pair of cycles C_i' and C_j' with $\ell(C_i') = \ell(C_i) + 1$ and $\ell(C_j') = \ell(C_j) - 1$. It is irrelevant here whether the graph actually supports such a decomposition: What is important is that any transformation from one cycle partition to another that decreases cycle length variation (and, necessarily, retains average cycle length K/N) can be accomplished through a sequence of *virtual* transformations of this form (one can always embed G into a larger graph in which this can be done in a concrete way).

To see that such a transformation improves rate of reward, one just notes that the rate along any closed cycle is $1 - \beta^{\ell(C)}$. For convenience, set $\ell_i = \ell(C_i)$ and $\ell_i' = \ell(C_i')$. Then the difference between the rate of reward along the original partition, versus along the new (perhaps virtual) partition, is

$$\begin{aligned}
\left[(1-\beta^{\ell_i'}) + (1-\beta^{\ell_j'})\right] - \left[(1-\beta^{\ell_i}) + (1-\beta^{\ell_j})\right] &= \left[\beta^{\ell_i} + \beta^{\ell_j}\right] - \left[\beta^{\ell_i+1} + \beta^{\ell_j-1}\right] \\
&= \beta^{\ell_i}(1-\beta) + \beta^{\ell_j}(1-\frac{1}{\beta}) \\
&= \beta^{\ell_i}\big((1-\beta) + \beta^{\ell_j-\ell_i-1}(\beta-1)\big) \\
&= (1-\beta)\beta^{\ell_i}\big(1-\beta^{\ell_j-\ell_i-1}\big) > 0
\end{aligned}$$

Figure 8: One agent following with delay

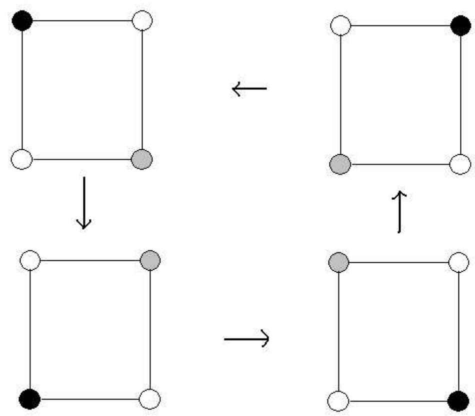
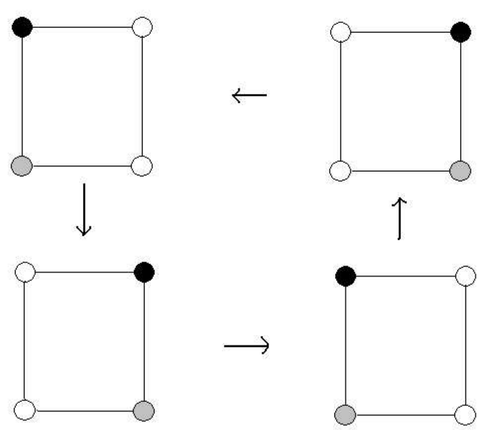


Figure 9: Parallel agents



since $\ell_j - \ell_i > 0$ and $0 < \beta < 1$ This proves the Theorem. ∎

5.2 Reward and capacity

5.2.1 Single agent grids

Suppose that a lone agent inhabits a graph G with the reward allocation rule $r(t) = 1 - \beta^t$ as before. Recall that G is Hamiltonian if it possesses a cycle (called a Hamiltonian cycle) containing every node. It is not difficult to prove the following:

Proposition 2 *If G is Hamilton then the rate of reward collection for $r(t)$ and for a single agent inhabiting G is maximized by traversing a Hamiltonian cycle.*

Once one sees that an agent cannot improve rate of reward by resting, the result follows easily from convexity as before. In the case of a $K \times K$ grid, the rate of reward then is $1 - \beta^{K^2}$ which tends to unit rate as K tends to infinity.

In the case of a non-Hamiltonian graph, a periodic agent (*i.e.* one that visits the same site every P time steps) collects reward at an average rate

$$R = \frac{1}{P} \sum_{steps} n(t)(1 - \beta^t)$$

in which $n(t)$ denotes the number of nodes that were last traversed t steps ago. Several nodes may be traversed more than or less than once during the circuit. In view of convexity, this rate will be largest when the large values of $n(t)$ are those close to the average return time. In this case R is also optimized when the *average* return time is as large as possible.

5.2.2 N agent grids and capacity

Theorem 1 implies that for our reward function $r(t) = 1 - \beta^t$, the optimal rate of collection by a system of N agents, where N divides the number of sites K, is at most $N(1 - \beta^{K/N})$. For any fixed $\beta \in (0, 1)$, this expression is increasing in N. Ignoring for the moment the fact that it may not be possible to put the N agents on disjoint cycles of length K/N, one concludes that the rate of reward should scale essentially linearly in N, at least for $N \leq K/2$.

When $N > K/2$ congestion plays a role: reward can only accumulate on sites that are open. But fewer than half the sides are open when $N > K/2$. With this in mind, it makes sense to define the capacity of a graph having K nodes and with reward $r(t) = r_\beta(t)$ to be $\text{Cap}_\beta(G) = \frac{K}{2}(1 - \beta^2)$. Then $\text{Cap}_\beta(G)$ provides a theoretical limit for the amount of reward that can be collected on the grid, on average per time step, for any system of agents.

Now that we have a reasonable definition of capacity for a Ping II grid, the next question is whether the decision making process with which the individual agents are endowed can and will result in coordinated activity – that is, aggregate behavior earning reward close to capacity. Because of the numerous sources of nonlinearity, not to mention boundary effects of the grid,

6 Simulation Observations

6.1 Risk factors and cumulative reward

Figures 10 through 12 show normalized cumulative system reward as it depends on the risk factor (here in the form of a multiple of the risk minus reward). Each symbol plotted represents a normalized production (*i.e.* cumulative reward) after 10^5 time steps. The cumulative reward is averaged over 5 random seeds, and divided by the number of agents. These individual averages are then normalized relative to the other risk

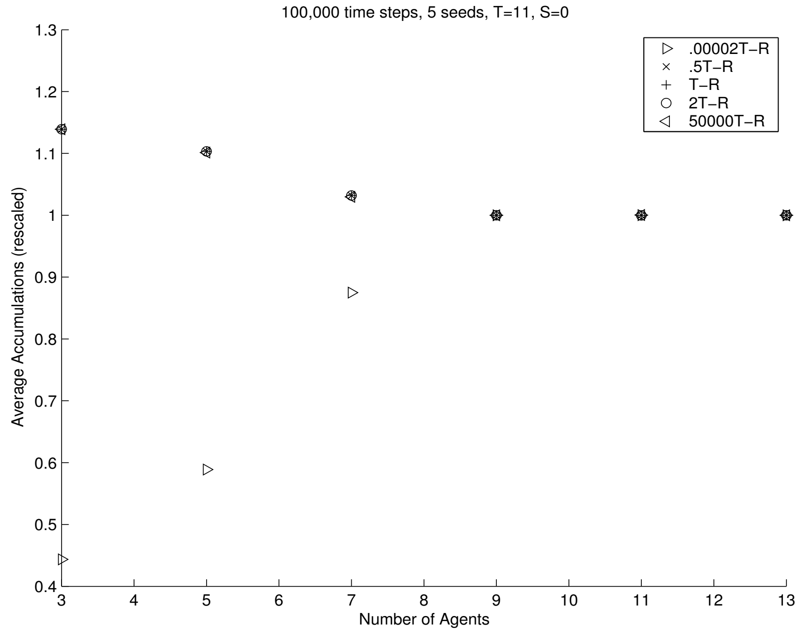


Figure 10: Normalized reward depending on relative risk and number of agents for threshold $T = 11$ and strength $S = 0$

factors by dividing out by the sum of the averages for the various risk factors (then multiplying by 5 for the 5 different risk factors). These normalized production values are plotted versus the number of agents. The grid size is fixed at 4×4 so the critical number of agents is $K^2/2 = 8$. In Figure 10 one further fixes the threshold $T = 11$ and strength $S = 0$. In the congested regime ($N \in \{9, 11, 13\}$, accounting for or ignoring risk has little effect. In the uncongested regime ($N = 3, 5, 7$), however, risk averse systems perform poorly compared to risk ignoring ones. Since $S = 0$ and, hence, no delays ever occur due to pings, fear of retribution in risk averse agents is unwarranted. Since weights are only updated when pinging occurs, the negative effect of risk aversion is completely due to random initialization of weights.

In Figure 11, on the other hand, $T = 11$ still but $S = 10$ and plenty of pinging occurs. It is still true in this case that risk aversion has little effect in the highly congested regime ($N \in \{11, 13\}$) and risk aversion still has negative consequences in the subcritical regime ($N \in \{3, 5, 7\}$). However, risk aversion leads to improved performance in the supercritical case $N = 9$. It is also worth mentioning that ignoring risk seems to lead to suboptimal performance when $N = 7$. One might be inclined to attribute this reduced performance to random effects (only 5 seeds), except that much the same behavior persists in Figure 12 where $S = 20$. Thus, one sees the importance of balancing greed and risk aversion on near critically congested grids.

In Figures $13 - 15$ we look at the total production of the system (averaged over 5 seeds) for different risk factors and numbers of agents, again for threshold fixed at $T = 11$ and $S = 0$, $S = 10$ and $S = 20$ respectively.

For comparison, Figures 16 and 17 plots total productions when the strength is fixed at $S = 10$ for values $T = 1$ and $T = 6$ of threshold.

In all of these figures we see that the dependence of production on risk factor goes away in the highly congested regime. On the other hand, in the uncongested regime we see that extreme risk aversion performs poorly but that mild risk aversion can improve production, particularly near the critical regime. As before, extreme risk aversion can actually lead to better production in the supercritical region. Figure 17 is partic-

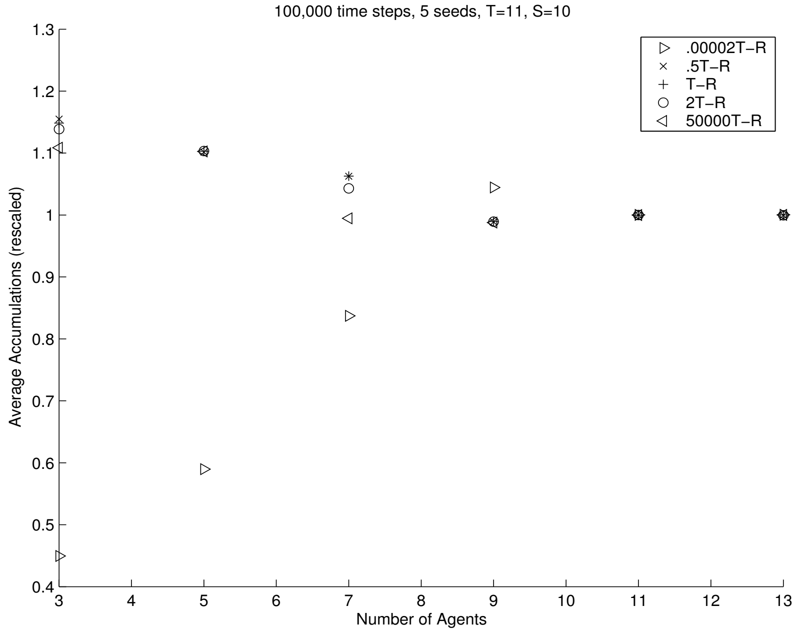


Figure 11: Normalized reward depending on relative risk and number of agents for threshold $T = 11$ and strength $S = 10$

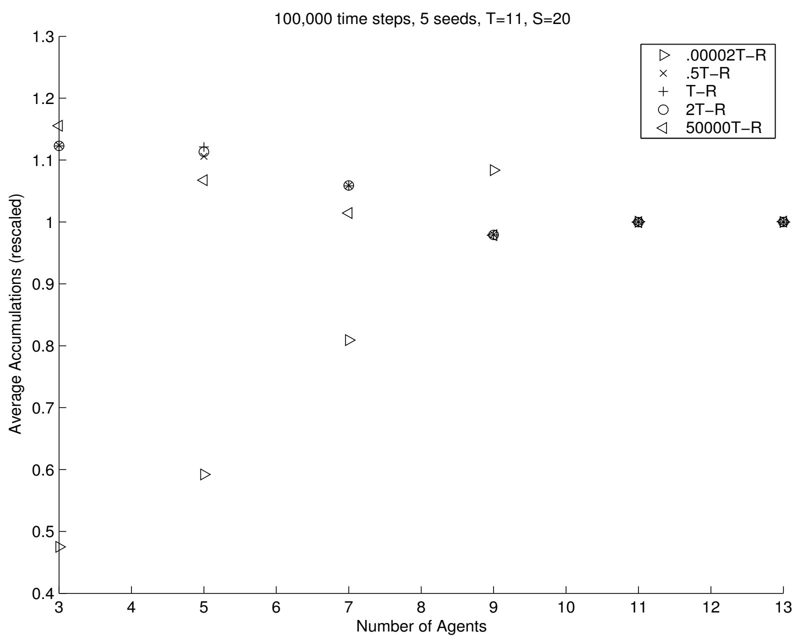


Figure 12: Normalized reward depending on relative risk and number of agents for threshold $T = 11$ and strength $S = 20$

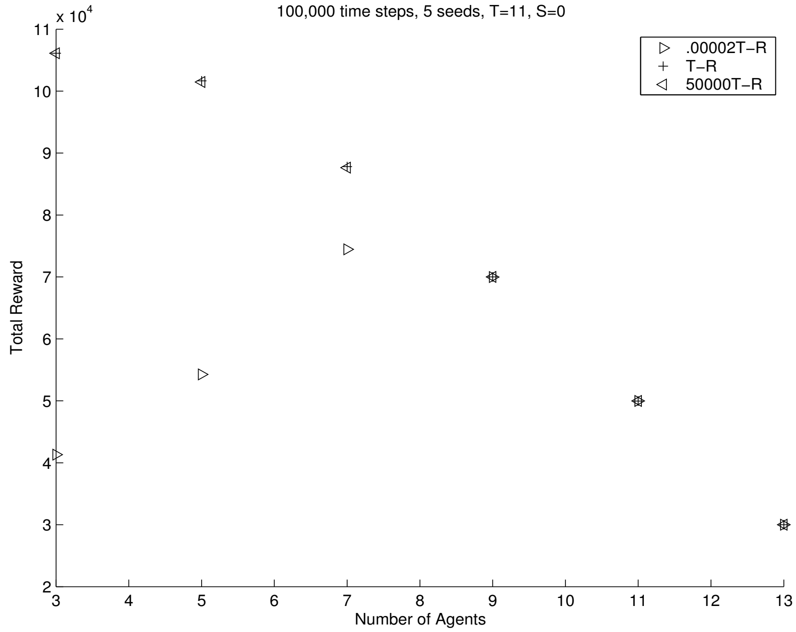


Figure 13: Total production depending on relative risk and number of agents for threshold $T = 11$ and strength $S = 0$

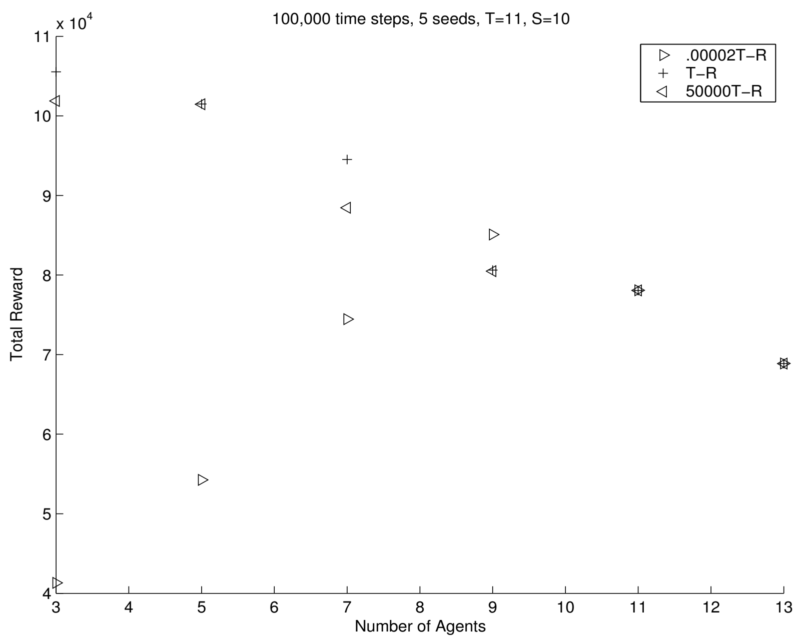


Figure 14: Total production depending on relative risk and number of agents for threshold $T = 11$ and strength $S = 10$

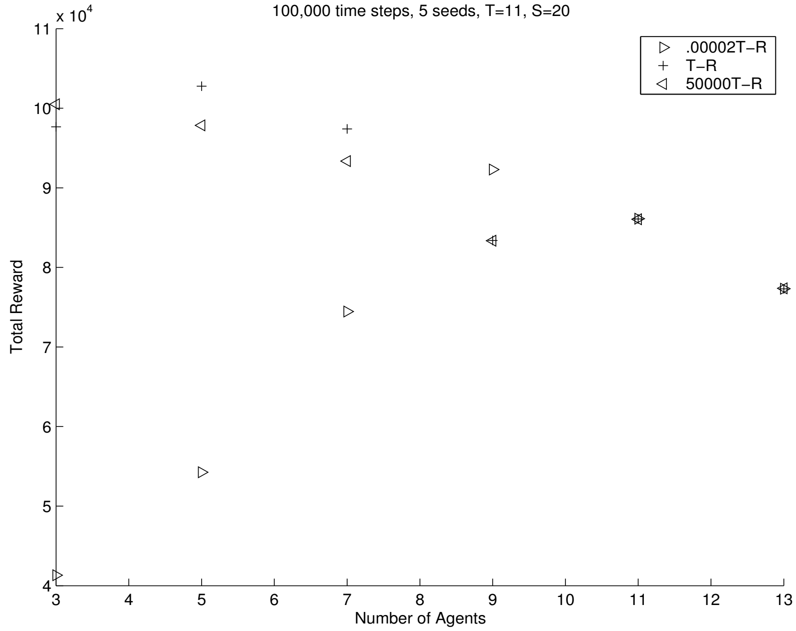


Figure 15: Total production depending on relative risk and number of agents for threshold $T = 11$ and strength $S = 20$

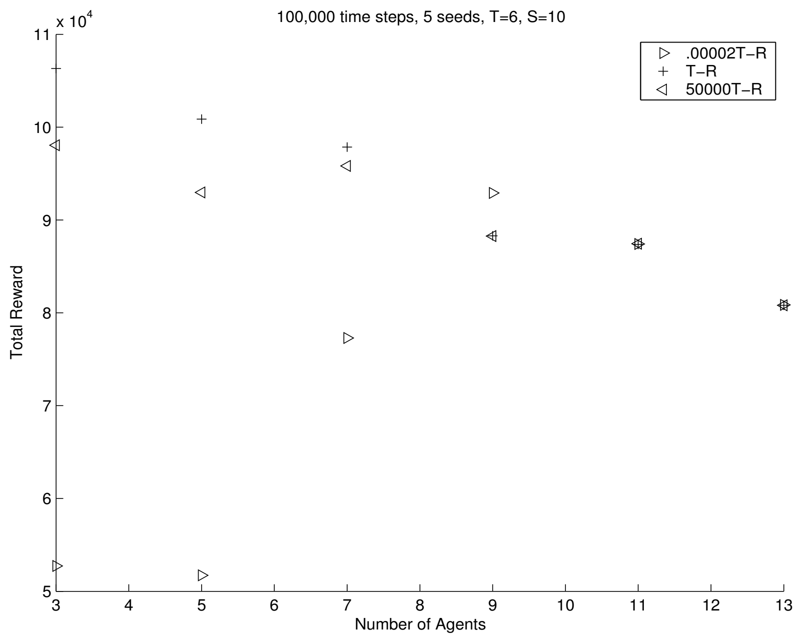


Figure 16: Total production depending on relative risk and number of agents for threshold $T = 6$ and strength $S = 10$

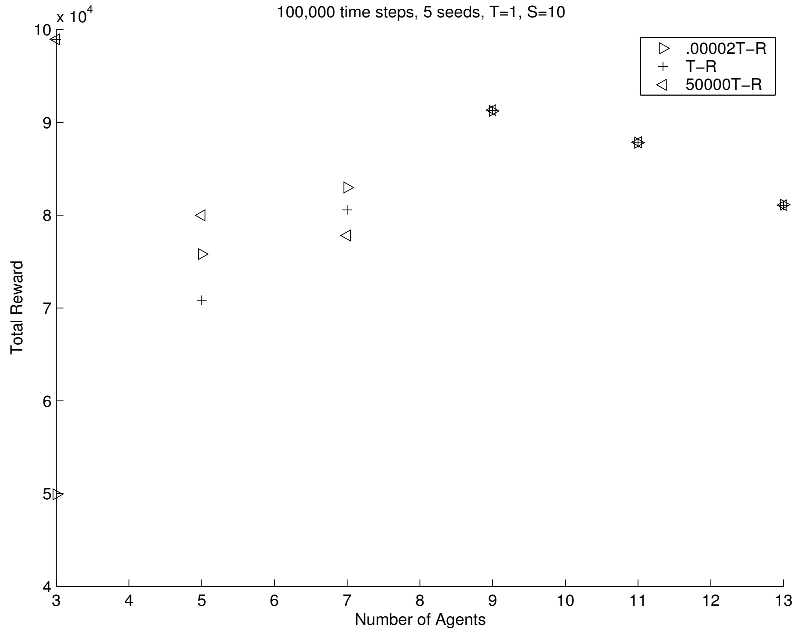


Figure 17: Total production depending on relative risk and number of agents for threshold $T = 1$ and strength $S = 10$

ularly telling: in this case agents have a high strength but very low threshold, so pinging results from little provocation. Here, in the highly uncongested region agents can avoid one another so risk aversion can still result in poor production. However, benefits of risk aversion are now seen in the subcritical region.

6.2 Geometry in small systems

In a 4×4 grid, 12 of 16 nodes are on the boundary and 4 of these are corners. This means that directions of movement on such a grid are highly constrained. Because fewer options for movement available, small systems can get into non-optimal limit cycles quickly. In many cases, unlucky agents are 'pinned' in a corner and have little opportunity to collect reward. In larger systems reward tends to be distributed over agents more homogeneously. Thus, boundary effects can lie at the root of subcapacitary performance, particularly in small systems. We are not yet reporting simulation results on this phenomenon due to increased computation time needed to show that this problem is less prevalent in larger systems.

6.3 Periodicity of position

Systems that are either highly constrained (congested) tend to become locked into periodic patterns and this limiting behavior tends to be induced by risk aversion. But production on these patterns can be suboptimal when agent paths have frequent self-intersections. This suggests considering a method to pass from shorter cycles to longer cycles.

7 Further Theoretical Developments Specific to Ping II

In Ping II there is minimal information sharing. At the same time, global knowledge of configurations plays an intrinsic role in computing risk, but does not play any role in encouraging individual exploration. Since risk aversion does not always automatically or rapidly lead to optimized system performance. There are two options to try to enhance or speed up system performance. One is to endow agents with improved rational abilities; a second is to allow agents to communicate with one another (or both).

7.1 Toward a reflective agent decision process in Ping II

One shortcoming of the risk-reward calculator approach is that it does not encourage an agent to take account of its own situation. An agent for whom the calculator is producing better than expected reward has no reason to alter its behavior. However, an agent that finds itself collecting sub-par reward should alter its behavior. One way in which to do so is to attach a higher weight to reward than to risk.

Because Ping II so far models *coordination* only of reflexive agents and, in particular, ones with little awareness of self in relation to others, no such adjustments are made in the current version. Insofar as one desires to model greed as opposed to benevolence, it should be assumed that any additional mechanism to improve global performance should also increase the utility of agents employing that mechanism. In view of Ferber's principles, the appropriate mechanism is communication via information attached to tags. In this sense one views the grid as a marketplace and the communication amounts, essentially, to an offer for exchange of property that is under an agent's control. In Ping II, what is *believed* to be owned is a collection of sites that is decomposed into cycles. The issue is not merely the *number* of sites owned: their connectivity structure is crucial in determining their aggregate value. As such, an agent can benefit from giving up control of some subset of nodes in exchange for other nodes that lead to longer cycles of ownership. One benefit of the extended rationality attached to agents who can negotiate accordingly is that even agents who are performing better average can still be motivated by greed to leave negotiation information.

7.2 Information sharing and cycle interchange

7.2.1 Agent awareness

There are many basic issues concerning an agent's awareness that affect its ability to make informed choices of movements. Some basic questions include: does each agent know the full capacity of the system, does each agent know the topological structure of the graph that it inhabits, and what awareness does each agent have of its immediate neighbors and neighbors up to a fixed distance at any particular time? For the sake of describing further a mechanism for negotiation, let us pretend for the moment that all agents are travelling on periodic (limit) cycles of ownership. In this case, it is reasonable to posit that an agent will have full knowledge of its periodic movement, that is, of all the sites that it visited and in what order. Moreover, an agent should reasonably be fully aware of those agents that have left tags on its path. What is less clear, but critical from the point of view of negotiating for territory, is whether an agent should retain full memory of agents in its neighborhoods at every stage of its traversal.

7.2.2 Cycle interchange

Suppose that two agents share a common node or are neighbors at some stage in their respective cycles. It is possible that both agents in the pair can improve their collection of reward by 'negotiating cycles'. The following example illustrates typical limit behavior for a pair of agents on a 3×3 grid trained by the risk calculator. In Figure 18 the darker circles denote nodes owned by Agent A_1 and the lighter those owned by agent A_2. If the tag decay factor is β then the average reward collected by agent A_2 is $1-\beta^4$ since A_1 moves in a cycle of length 4. On the other hand, the average reward collected by A_1 is $1 - \frac{1}{4}\sum_{k=1}^{4} \beta^{2k}$.

In Figure 19 the two agents have negotiated to share the middle row (black nodes). In this case each agent moves on a cycle of length 6. Because of the relative positioning, each agent now collects reward at an

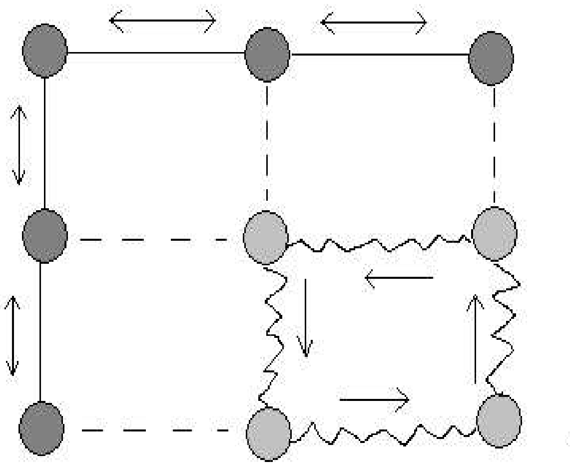


Figure 18: Typical limit paths of agents A_1 and A_2

average rate of $1 - \frac{1}{2}\beta^3 - \frac{1}{2}\beta^6$. For values of β between approximately $\beta = 0.619$ and $\beta = 0.667$, the average rewards collected under the movement of Figure 19 are larger than those collected under the movement of Figure 18 for both agents, *provided* the two agents do not ping one another when they are neighbors. The movement in Figure 19 dictates that the two agents will frustrate one another (they share three nodes in common) while they will be neighbors when one of the agents is in the middle node. Therefore, the only way that both agents can benefit is by entering an agreement that they will not ping one another while moving as in Figure 19. It is worth noting that any relative positioning of A_1 and A_2 along paths in Figure 19 such that A_1 and A_2 are never neighbors implies decreased average reward from the original paths for at least one of A_1, A_2.

In order to enter a negotiation an agent must be able to deduce a benefit in entering a pact. In this example, for the given range of the parameter β, both A_1 and A_2 must be able to discern that they will benefit from making the indicated adjustments. This is possible if both agents have an awareness of all neighborhoods in their respective paths as, in this case, every node on each new agent path is a neighbor of one on the old agent path. This is not the only possible manner in which increased rationality can lead to mutually beneficial behavior but it indicates one reasonable mechanism for such benefit.

8 Further Modelling Concerns: Disruption of Coordination

In Section 2 we noted several forms of deviance from 'rationale' behavior that pose difficulties in resolving asymmetric threats. A particular theoretical concern that has come to light through work of J.M. Carlson and J. Doyle and others, (*e.g.* [3]) is that systems based on dynamical rules that are specifically designed to optimize certain types of behavior can often exhibit unintended vulnerabilities.

Although, as described, Ping II does not readily reflect a social dilemma. Moreover, the reflexive agent behavior does not lend itself easily to successful contrarian tactics. That is, insofar as deviance amounts to *behavior contrary to the accepted rules*, in the context of Ping II deviance can only mean (i) pinging under inappropriate circumstances and or (ii) moving counter to direction indicated by the risk calculator.

Once cooperative behavior is proposed to improve production – cycle exchange being one possibility –

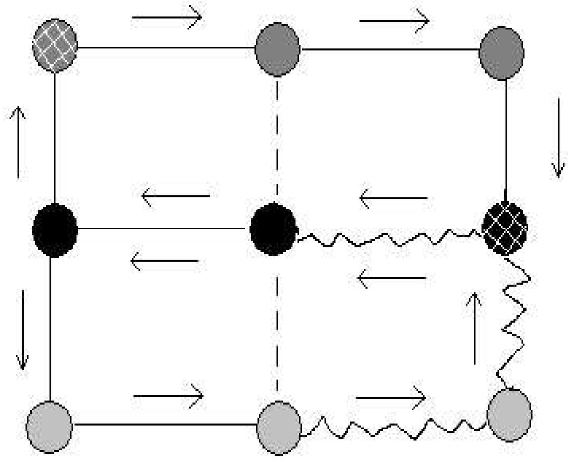


Figure 19: Negotiated paths for A_1 and A_2

deviant behavior also has more possibilities. For example, the proposed method of cycle interchange is based on agent awareness of its neighbors. However, one does not assume knowledge of nodes at greater distance. A deviant agent can take advantage, for example, by negotiating cycle exchange that would place another agent at higher risk of coming into contact with a frustrated agent (for example by passing false information about second neighbors). A sequence of such exchanges could substantially diminish overall production.

In the original project description for this contract, we proposed to analyze deviant behavior as part of possible applications of the simulation/modelling effort in the realm of homeland security. The original context of the proposal was Ping like *hide and seek networks*. Similar notions have recently gained some attention in the computer network security realm in the context of making local area networks moving targets for hackers (*e.g.*, `http://www.newscientist.com/news/news.jsp?id=ns9999776`). This is done by means of randomizing and encrypting the local part of a computer's IP address – roughly analogous to the manner in which a terrorist cell might use mutable digital signatures to enable communication but, at the same time, avoid detection.

To model such a situation an agent should have a public identity (agent number in Ping II) but also a *group identity* that can remain hidden to others outside the group. Group identity is an important aspect of social network modelling, as is the case in Choi's work [5] on local cooperation/defection outlined below. In that context, defection can be phrased in economic terms just as cooperation. In the present context we regard deviance as purposeful disruption of coordination. As such, it should also have attached some form of virtual reward: namely the amount of reward that is prevented from being attained by means of deviant behavior. As such, we will consider possible mechanisms for the emergence of *cells of deviance*, just as one does so for coordination, in future iterations of Ping that are optimized via cooperation.

9 Further General Theoretical Issues

The Ping II model for coordination of activity in a social network bears important similarities, not only to other mobile social networks in the spirit of Axtell and Epstein's Sugarscape [1] but also of more generic games in which cooperation or lack thereof is the principle determinant of system performance. Rigorous evaluation of learning methods in the context of such games is a fairly recent development (see Holland [14] who suggests that things progressed little beyond Samuel's 1959 checkersplayer [20].

On the other hand, Ping II exhibits several features of a mathematical *landscape.* That is, a triple consisting of a space of *configurations*, a *fitness* function that evaluates those configurations and, for each configuration a set of neighboring, or accessible configurations under allowable *moves.* Insofar as configurations can be identified with nodes of a graph and moves with its edges, landscapes are amenable to spectral graph decompositions. These decompositions can yield important information about the basic *modes* under which the fitness function can be decomposed.

The sense in which these tools – learning theory of social games on the one hand and spectral theory of landscapes on the other – can be brought to bear on Ping II and foreseeable derivatives of it, will be discussed in what follows.

9.1 Learning theory of social games and networks

Learning theory in social games has developed largely in the context of particular examples. A recent exception to this is work of Foster and Young [10] addressing N-player matrix-payoff games. What is common to all of the approaches that will be outlined below, though – and not included in the current setup of Ping II – is that performance can be improved *on average* when players employ a strategy based on reasonable probabilistic assumptions. Learning then is essentially a matter of learning what proportionality or frequency should be assigned to a given strategies of play. Moreover, it is important to note that optimal global performance is typically realized when there is no arbitrage strategy. This gives some sense of the ability to take advantage of an otherwise optimized system (*cf.* [3]) by means of a deviant strategy.

9.1.1 Learning of coordinated movement

Coordination was studied in the context of mobile social networks by D. Vengerov, H. Berenji, and A. Vengerov [24]. As is the case of Ping II, reward is allocated at sites on a grid and collected by agents. In their system, however, reward is allocated probabilistically. Moreover, neighborhoods are taken in the sense of Moore – not von Neumann as in Ping II.

The agent simulation of [24] places models in a 2-D tileworld. Its goal is to have agents learn a topology-dependent coordination strategy. Randomness is introduced by allowing demand sources to disappear and reappear at a randomly chosen site with given fixed probability p while newly appearing reward sources have a reward value R_j uniformly distributed between 0 and R. This redistribution of reward (i) is not overly capricious from the agent's viewpoint provided p/R is small enough and (ii) allows one to model uncertainty in localization of reward, to some extent.

The sources together form a grid potential whose value at node i is $P_i = \sum_j P_{ij}$ where $P_{ij} = \frac{R_j}{1+\delta_{ij}^2}$ with δ_{ij} the distance of the j-th source from node i. Reward is also updated through extraction: an agent at location i extracts a reward of P_{ij} from source j.

Ideas from game theory come into play here because extraction by agents can rapidly deplete reward, while agents cannot necessarily guess one another's intentions regarding where to move next. Consequently, agents tend toward regions having high expected resource potential but low expected agent demand. The decision to move in a direction indicating high reward minus demand is similar, though also different in important ways, to our calculation of reward minus normalized risk. For example, unlike the deterministic neural net employed in Ping II, Vengerov *et. al.* employ *fuzzy Q-learning* to update strategies. while agents use *fuzzy Q-learning* to decide on where to move. Fuzzification lumps risk/reward values into categories and may be more appropriate than use of precise values as in Ping II.

The simulation results reported by Vengerov *et. al.* were based on a 20×20 grid populated by 5 agents and 10 sources of reward. In comparison to Ping II, this would correspond to an uncongested network since agents can easily avoid getting in one another's way, provided the reward sources are well distributed. The main conclusion was that coordinating agents learned to prefer locations having smaller agent demand, that is, to avoid one another. A gain of 50 to 100 percent in performance (average reward) was reported when fuzzy Q-learning was employed in contrast to random movement. Moreover and importantly, agents could make global evaluations of potentials and not just the nearby potential field: agents with limited sensory radius were reported to perform worse than randomly acting agents.

9.1.2 Coordination versus cooperation

In Ping II, the level of coordination is measured by the rate at which the system can collect reward and the global neural net risk-reward calculator provides one possible mechanism for coordination. The question to what extent can this calculator lead to optimized activity is a *constraint alignment* problem, one that we have framed in terms of structure of graph partitions.

As we have observed, because agents are constrained to move locally, reward can be distributed unevenly in the limit just because of topological obstructions, particularly in small systems. We have already discussed a possible cooperation mechanism for better performance among intelligent agents. However, in the base case of Ping II, agents do not cooperate and the basic issue remains: how can one coordinate activity among fundamentally uncooperative agents?

9.1.3 Probabilistic learning in minority games

Minority games provide a fairly simple model for the paradoxical question of how one can get agents to coordinate in a noncooperative setting. The standard minority game is a 'simple' model for more complex coordination problems such as Arthur's El Farol bar problem. One has an odd number N of players. An agent receives an equal reward (say one point) if it is in the minority among the players in casting a *vote* of $\epsilon \in \{0, 1\}$. An agent who votes with the majority earns no reward. Agents are thus rewarded for *dissenting*. Optimum reward occurs at each of the $\binom{N}{(N-1)/2}$ Nash equilibria for which the minority consists of $(N-1)/2$ agents.

A *strategy* in this game is a mapping that assigns a play on the next step based on the history of all previous winning outcomes. A typical strategy here might be simply to vote contrary to historical average. Thus if, up to time T, $\epsilon = 1$ has more victories than $\epsilon = 0$, an agent employing this strategy will vote $\epsilon(T + 1) = 0$. While such a strategy might seem reasonable, it can be a real problem if each agent chooses the same strategy: then every agent ends up voting the same way and $M = 0$. This tragedy results from inability to model competing agents. Supposing that agents have no predisposition, can an agent learn an effective strategy? This turns out to be an ill-posed question as will be discussed below. The main problem is that an agent must somehow account for the ability of other agents to react to its strategy. This makes playing any particular deterministic strategy a bad idea.

Thus it makes sense to consider completely random behavior as a *base case*. In that case the *variance from optimum* is $V = E(((N - 1)/2 - M)^2)$ where M is the number of minority agents and the expectation E is determined by the binomial distribution: Since the minority has size M with probability $\binom{N}{M}/2^N$, one has $V = \frac{1}{2^N}\sum_{M=0}^{(N-1)/2}((N - 1)/2 - M)^2\binom{N}{M}$. For example, when $N = 11$, $V \approx 3.56$. The point of learning, clearly, is to lower this variance or *volatility* by making the minority larger with a higher probability.

It turns out to be fruitful to define a space of strategies, stratified by memory length. A strategy s has *memory* m if it depends only on the prior m outcomes. Then s is a deterministic mapping $s : \{0, 1\}^m \to \{0, 1\}$. A mixed strategy of length m is a probabilistic strategy that assigns probabilities of play to some subset of strategies of length m.

Bottazzi, Devetag and Dosi [4] considered a method for learning a mixed strategy, that is, learning the most effective probabilities to assign for play among a fixed collection of deterministic strategies. An agent is said to have memory of length m when these strategies do. In each round, agents are informed whether 0 or 1 is the minority outcome, but not which agents are in the minority. The parameters m, s govern the agents' rationality, although the logic for assigning probabilities of play certainly is a factor as well.

Importantly, different agents are assigned different sets of strategies from which to choose. An agent can only change the frequency with which the strategies are played - not the strategies themselves. These frequencies are updated based on *virtual reward*. That is, the agents assign probabilities of play based on the frequency with which each strategy, if played, would have won. Precisely, if $q_i(t)$ is the total number of points that strategy i would have produced up to time t then an agent chooses among its strategies for next play according to

$$p_i(t) = \frac{e^{\beta q_i(t)}}{\sum_j e^{\beta q_j(t)}}$$

in which the sum on j goes over all strategies possessed by the player. The factor β can be thought of as a learning rate. Phrased thus, one then studies how system payoff depends on memory length.

A reasonable measure of success is the temporal mean squared deviation:

$$\sigma = \frac{1}{T} \sum_{\tau=0}^{T} (N_0(\tau) - \frac{N}{2})^2$$

where $N_0(\tau)$ is the number of agents playing zero at time τ. The main question is: how does this quantity depend on the learning rule?

To begin to answer this question one defines a parameter $z = 2^m/N$ that provides a measure of the density of agents in the strategy space. There are 2^{2^m} strategies of length m. When z is large then there are too many strategies for the system to *self-organize*. Conversely, $z << 1$ corresponds to an *inefficient regime* in which agents densely populate the strategy space. Actions are strongly correlated and performance is poor due to this *crowd effect*.

In between lies a *stable regime*, $z \approx 1$. Here coordination produces better than random performance: there is enough differentiation among strategies to avoid crowding but sufficient distribution of agents over strategies to avoid random behavior. This is analogous to Ping II in which the density of agents, not in strategy space but on the physical grid, has a substantial impact on the efficacy of learning.

The general results indicated by Bottazzi *et. al.* suggest that, for a fixed number of agents N, the volatility first decreases as $m \sim z$ increases from the inefficient regime toward the critical memory length, then increases as $m \sim z$ passes into the random regime. Though the minority game is quite different from Ping II, this observation is, nevertheless, consistent with what we have observed about the relationship between density of agents on a grid (analogous to density of agents in strategy space) and efficacy of risk evaluation. The learning rate β also interacts with memory length m in determining long range volatility. In particular, β plays a stronger role in the inefficient regime: smaller β yields (albeit more slowly) smaller long range volatility in this case but has little or no bearing on the case $m \sim z$ large. We have yet to evaluate the effect of learning rate (our factor Δx) on performance in Ping II.

The parameter β can be viewed as a willingness to choose. For large β, agents are sensitive to small changes in virtual reward while, when β is small, they are more stable. Suppose that all but one of the agents are playing, each according to their own specific strategies. Suppose that the exceptional agent has knowledge of a longer history. One asks: if the exceptional agent is playing the best possible strategy (then referred to as an *arbitrager*), how well will it perform relative to the other agents? Curiously, β plays a significant role here: if β (the learning rate for the other agents) is large then arbitrage can yield significant gain whereas, if β is small then there is little *arbitrage opportunity*.

An interesting alternative to the N-player minority game was proposed by Galstyan and Lerman [12]. Non-stationarity is introduced by making the size of the winning 'minority' dependent on a function $\eta(t) \in [0, 1]$. Precisely, the winning choice is "1" if $A(t)$, the sum of those who 'played' "1" is at most $N\eta(t)$. The winning choice is "0" otherwise. Since 'minority' depends on t, it makes sense to define volatility here as

$$\sigma^2(T) = \frac{1}{T} \sum_{t=1}^{T} (A(t) - N\eta(t))^2.$$

Strategies are defined slightly differently from the standard way. Namely, rather than taking into account history of the game up to time t to make a choice of play, strategy at time $t+1$ is defined in a Markov sense, solely as a function of the play of a fixed set of K other players at time t. As with Bottazzi *et. al.*, however, agents are randomly initialized with a fixed set of strategies that are played with probability depending on their history of virtual success. Several curious results are reported by Galstyan and Lerman. First, it is reported that the system adapts to the nonstationary 'capacity' best when $K = 2$, at least provided η does not fluctuate too rapidly. For $K = 1$ there is reaction to the nonstationary resource but volatility is large, while for $K > 2$, response to environmental dynamics deteriorates as K increases. This phenomenon is independent of the number of agents. It seems significant here that, in all cases, mixed strategies were learned from *two* randomly signed initial strategies: it seems possible that this is why $K = 2$ leads to coordination, since $K > 2$ agents might be providing *inconsistent information* from the perspective of the agent updating its strategy probabilities. The authors do not address this question. They do, however, compare their results with the standard memory length version of strategy. Simulation results indicate that utilization of resources in the non-stationary case is better when mixed neighbor strategies are learned as opposed to mixed memory strategies, while variation among agent performance is also smaller in the neighbor-learning case.

9.1.4 Act locally, learn globally

The global risk evaluator in Ping II was designed on intuitive grounds. Recent work of Choi [5] indicates the appropriateness of employing global information, at least in *cooperation* problems. Choi employed genetic learning in the context of a social dilemma game in which cooperation and coordination, in a sense, amount to the same thing. Here, one starts with a collection of N agents divided into T groups each of size $n = N/T$. At each round one assigns a cost c to cooperation and a benefit b to belonging to a group. Denoting by p_j the fraction of cooperators in group j, the payoffs for cooperators and defectors in the j-th group are then

$$\begin{aligned} \pi_j^C &= bp_j - c \\ \pi_j^D &= bp_j \end{aligned}$$

Defectors always have better payoff than cooperators within a group, but everyone in the group has low payoff if there are few cooperators.

Modelling cooperation or defection with the binary variable $\{0, 1\}$, a *strategy* is a mapping from $\{0, 1\} \times \{1, \ldots, n\} \rightarrow \{0, 1\}$ providing a decision to cooperate or defect based on (i) whether the given agent cooperated in the previous step and (ii) the total number of cooperators in the previous step.

Having a whole collection of groups allows agents to learn globally by following strategies of players in other groups. This helps to avoid dilemmas that arise when an agent must base its actions solely on expected play of others in the group. Choi has shown that, under a suitable, simple *genetic* learning rule, systems in which agents choose *teachers* globally result in better global payoff than ones in which agents are constrained to learn locally. It is not clear that this basic point of view has ramifications for Ping II, although results of Vengerov *et. al.* suggest that, in uncongested grids, ability to take into account (spatially) global state information can result in better coordination.

9.1.5 Learning Nash equilibria

In Ping II, agents act asynchronously (according to a queue). Even so, an important consideration of simultaneous play games is still relevant, namely: what are the intentions of agents with whom a given agent must interact? This issue arises in several scenarios just considered but the results are all based on simulations. When coordination is phrased in terms of Nash equilibria, Foster and Young [10], [9] have established some important basic principles concerning what can be achieved through probabilistic learning. In particular, in the context of multiplayer games, hypothesis testing can yield *success with high probability*, provided the test is sufficiently powerful and, crucially, *no hypothesis is rejected outright*. Moreover, Foster and Young show that deterministic learning is doomed to failure for much the same logical paradox that arises in iterated prisoner's dilemma, namely the conflict that arises in predicting play of other agents based on their history and, at the same time, their reaction to the given agent's history. This dilemma confronts Ping II agents, at least in a weak sense, since an agent's action can have nontrivial unintended consequences.

Nonetheless, we will consider the consequences of probabilistic player in future iterations of Ping II. As such, hypothesis testing of neighbor agent intentions can have a significant effect on coordination.

9.2 Landscapes and spectral theory

Ping II exhibits salient features of a mathematical *landscape* (*e.g.*, [23]) on several levels. A landscape is a triple consisting of (1) a space of configurations, (2) a collection of rules, deterministic or probabilistic, for transition from one configuration to the next and (3) a *fitness* function that evaluates each configuration. Landscape theory can be viewed as an approach to explicate the decomposition structure of a landscape, thereby adding valuable insight to the behavior of a landscape viewed as a complex system. To make the theoretical tools available, however, the system needs to be presented as a landscape in appropriate terms.

At the most basic level, a Ping II system can be viewed as a landscape in which the configuration consists either simply as the arrangement of agents on the grid or, in a more complex form, as this arrangement together with the set of tags. In the latter form the rule for transition among configurations is determined by the neural net or some variation thereof. It is tempting to define the fitness of a configuration in this form as the reward collected by agents in this configuration. However, this approach does not shed light on the evolution of the system. Specifically, there may be little or no correlation between fitness of a given configuration and of 'neighboring' ones. In the vernacular, the landscape may be very rough, reflecting the reflexive behavior of the agents. Roughly speaking, rough landscapes are difficult to decompose into components that enable predictions about behavior.

On the other hand, if one views configurations simply as arrangements of agents, the only reasonable way to define the fitness of a configuration is through an empirical time average that destroys important information about the particular conditions leading to high or low rewards.

When one wishes to investigate the consequences of reflective agent rules, however, it makes sense to define a type of hyper-landscape. In this approach, configurations are defined in terms of limit cycles determined by initial simulation conditions with reflexive agents. Fitness of such a configuration can be defined then in terms of the average reward collected over one system period. A rule for transitions among limit cycles is needed. One such rule can be defined in terms of cycle exchange as outlined above. Transition rules are necessarily more complex in the case of reflective agents. On the other hand, a good set of reflective rules will induce "hill-climbing" in the configuration space, thus leading the system to fitness peaks without "external intervention". General tools, particular from spectral landscape theory, can provide valuable information regarding not only what sorts of move rules to use in order to induce hill-climbing, but more fundamentally, which hills are good to climb.

9.2.1 Graph Laplacian and Schrödinger operators

Any landscape can be thought of as a graph in which the nodes are the configuration and the edges are those pairs of configurations that are joined by allowable moves. To think of moves in terms of flows on a graph, one introduces the graph Laplacian. One way of doing so (see Stadler, [21], [22], [23]) is to start with an orientation function that assigns an origin and terminus to each edge. Taking such an orientation into account one then defines $\nabla^+(u,e) = \mathrm{sgn}\,(u,e)\sqrt{a(u,e)}$ where $\mathrm{sgn}\,(u,e) = 1$ if u is the tip of e and -1 if u is the base, and 0 if e is not attached to u. If one also sets $\nabla f(e) = \sqrt{a(u,v)}[f(v) - f(u)]$ (here $e = (u,v)$ where the pair is ordered in the sense that u is the base and v is the tip). With this definition one sets $\Delta = -\nabla^+\nabla$. The spectrum of a landscape is the spectrum of its graph Laplacian Δ.

One of the main uses of spectral graph theory in this context is in providing a qualitative formulation of ruggedness of a landscape in terms of the relative concentration of fitness peaks – both spatially and in terms of amplitudes.

One important aspect of this analogue between the continuous and discrete settings is Courant's Nodal Domain Theorem. Let ψ_k be an eigenvector of H with eigenvalue λ_k. Here $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{|V|}$ so the eigenvalues are increasing and not necessarily distinct. Stadler does not say anything about the signs of the eigenvalues. One sets $m(k) = \min\{m : \lambda_m = \lambda_k\}$ and defines $M(k)$ similarly, with min replaced by max. Define a strong nodal domain of a wavefunction ψ to be a maximal connected component of $\Gamma_+(\psi)$ or $\Gamma_-(\psi)$, where $\Gamma_+(f) = \{u \in V : f(u) > 0\}$. A weak nodal domain is a maximal connected component of $\Gamma_+(\psi) \cup \Gamma_0(\psi)$ where the latter denotes the zero set. By connected here we mean connected by a path within the given set.

The nodal theorem says that there are at most $M(k)$ strong nodal domains and at most $m(k)$ weak nodal domains. Moreover, if ψ_k has $m(k) + \ell$ strong nodal domains then every vertex meets at least $\ell + 1$ of them. The last statement gives some intuitive feel of the landscape.

It is worth mentioning that the discrete nodal domain theorem also applies to any reasonable Schrödinger operator on a graph. Such an operator has the form $H = -\Delta + V$ in which Δ is the graph Laplacian and V is a *potential function* (*e.g.* a distribution of reward). In the Euclidean case, eigenfunctions of this operator correspond to *bound states* and the *uncertainty principle* provides estimates on the localization of such states. The correspondence between potential functions on a grid and corresponding landscape potentials will be investigated in future work.

9.2.2 Deceptiveness

One important aspect of landscapes is the complexity of the fitness function. In terms of spectral theory, this complexity can be formulated in terms of the rate of decay of coefficients in an eigenfunction expansion of the fitness function. Slow decay can imply the presence of fluctuations that cannot be determined by looking at the first few (*i.e.* dominant) terms of this expansion. Deceptiveness can also be phrased, roughly, in terms of amenability to study via a genetic algorithm (GA) – the first step in trying to find a fitness optimizing configuration. Here deception means, roughly, that a GA has a low probability of landing near a fitness optimum. The two notions of deception – in terms of spectral expansions and in terms of GAs – are closely related when the landscape has the specific form of a (generalized) Boolean hypergraph. Insofar as these notions of deception can conform to notions of effective deviant behavior as outlined in Section 2, it is worth studying deceptiveness further in the context of Ping landscapes.

Finally, insofar as the GA approach is related to multiscale sampling, joint work of the PI (Lakey) with J. Hogan [13] will be extended further in the context of defining wavelets on graph structures and establishing sampling expansions in terms of wavelets – which are best thought of here as computationally feasible approximate eigenfunctions of the Laplacian.

References

[1] J.M. Epstein and R. L. Axtell, "Growing Artificial Societies: Social Science from the Bottom Up," Brookings Institution & MIT Press, 1996.

[2] A. Bechara, A. Damasio and A.R. Damasio, *Emotion, decision making and the orbitofronral cortex*, Cerebral Cortex **10**, (2000), 295–307.

[3] J.M. Carlson and J. Doyle, *Highly optimized tolerance: robustness and design in complex systems*, Phys. Rev. Lett. **84** (2000), 2529-2532.

[4] G. Bottazzi, G. Devetag and G.Dosi, *Adaptive learning and emergent coordination in minority games*, Simulation Modeling Practice and Theory **10**, (2002), 321–347.

[5] J.-K. Choi, *Play locally, learn globally: the structural basis of cooperation*, Santa Fe Institute technical report, 2002.

[6] A. Damasio, "Descartes' Error: Emotion, Reason and the Human Brain," Quill: New York, NY, 1994.

[7] R. Fagin, J.H. Halpern, Y. Moses and M.Y. Vardi, "Reasoning about Knowledge," MIT Press: Cambridge, MA, 1995.

[8] J. Ferber, "Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence," Addison Wesley: Harlow, UK, 1999.

[9] D. Foster and H. P. Young, *On the impossibility of predicting the behavior of rational agents* PNAS **98** (2001), 12848–12853 `http://gosset.wharton.upenn.edu/ foster/research/`.

[10] D. Foster and H. P. Young, *Learning, Hypothesis Testing, and Nash Equilibrium*, Games and Economic Behavior **45** (2003), 73–96 `www.santafe.edu/sfi/publications/wpabstract/200302007`.

[11] D. Fudenberg and D. K. Levine, "The Theory of Learning in Games," MIT Press, 1998.

[12] A. Galstyan and K. Lerman, *Minority games and distributed coordination in non-stationary environments* in "Proc. International Joint Conf. on Neural Networks," May 2002, Hawaii.

[13] J.A. Hogan and J. Lakey, *Sampling and oversampling in shift-invariant and multiresolution spaces I: validation of sampling schemes*, Int. J. Wavelets Multiscale Inf. Proc., *to appear*.

[14] J. Holland, "Emergence: From Chaos to Order," Perseus Books, Cambridge Mass. 1998.

[15] G. Klein, "Sources of Power: How People Make Decisions," MIT Press: Cambridge, MA, 2001.

[16] J. Lakey, M. Coombs, K. Streander and C. Weaver, *On models for coordination activity and its disruption: interim progress report, 06/01/02– 12/31/02*.

[17] J.E. LeDoux, *Emotion: clues from the brain*, Annual Review of Psychology **46** (1995), 209-235.

[18] M.D. Lieberman, R. Gaunt, D.T. Gilbert and Y. Trope, *Reflection and reflexion: a social cognitive neuroscience approach to attributional inference* in "Advances in Experimental Social Psychology," M. Zanna (ed.), Academic Press: New York, NY, 2002.

[19] B. D. Perry, *The neurodevelopmental impact of violence in childhood* in "Textbook of Child and Adolescent Forensic Psychiatry," D. Schetky and E. Benedek (eds.), American Psychiatric Press:, Washington, DC, 2001, 221-238.

[20] A.L. Samuel, *Some studies in machine learning using the game of checkers* in "Computers and Thought," E.A. Feigenbaum and A. Feldman (eds.), McGraw-Hill, New York, 1963.

[21] P. Stadler, *Spectral landscape theory* in "Evolutionary Dynamics - Exploring the Interplay of Selection, Neutrality, Accident, and Function," J.P. Crutchfield and P. Schuster (eds.), Oxford University Press, 2002, 231–272.

[22] P. Stadler and C.R. Stephens, *Landscapes and effective fitness*, Comments Theor. Biol. **8** (2003), 389 – 431.

[23] P. Stadler, *Fitness landscapes* in "Biological Evolution and Statistical Physics," M. Lässig and A. Valleriani (eds.), Springer-Verlag, Berlin, 2002, 187–207.

[24] D. A. Vengerov, H. R. Berenji and A. B. Vengerov, *Adaptive Coordination Among Fuzzy Reinforcement Learning Agents Performing Distributed Dynamic Load Balancing*, in "Proceedings of the 11th IEEE International Conference on Fuzzy Systems," (2002), 179–184.

[25] D. Wolpert and Lawson, *Designing Agent Collectives for Systems with Markovian Dynamics*, NASA Ames Research Center Paper ID:546.

[26] D. Wolpert and K. Tumar An introduction to collective intelligence. (2000) `http://ic.arc.nasa.gov/projects/coin_pubs.html`.