Smooth Dynamics, Good Performance in Cognitive-Agent Congestion Problems

David Reitter, Penn State University
Paul Scerri, Carnegie Mellon University

Abstract

In a congestion game, individuals exhaust a common resource out of selfish behavior. In this scenario, drivers create traffic jams by choosing the shortest route according to their individual knowledge. They can avoid them by communicating their belief states about the traffic situation in real-time through a peer-to-peer network, assuming unlimited bandwidth. We study two potential, cognitively inspired models of human behavior: 1) categorization (quantized memorization and communication), which dampens communication and belief adoption, but leads to undesired oscillations and lower performance. 2) Instance-based blending with memory decay, which achieves good dynamics and near-optimal performance without the same bandwidth needs. We argue that this supports our hypothesis of co-adaptation of cognitive function and communicating communities.

Introduction

In many situations, crowds of interacting human individuals share resources such as food, roads, electricity, internet bandwidth, or airtime in a conversation. Similarly there are many interesting domains that require robots or agents to simultaneously learn to utilize common resources. When the actions of one agent impact the outcomes of another agent, individual learning often leads to complex system dynamics. A canonical example of this problem is cooperative path planning(Burgard, Moors, Fox, Simmons, & Thrun, 2000), where agents using the same routes negatively interfere with one other, but many other domains have been studied including soccer(Kalyanakrishnan, Liu, & Stone, 2007) and markets(Tesfatsion & Judd, 2006).

Computational-cognitive models of social behavior require a combination of cognitive architectures and multi-agent design and analytics. In this paper, we investigate the effect of memory decay and instance-based learning and decision-making a system of communicating, simulated individuals.

Based on the rational assumption that human cognitive function has adapted to its environment, we hypothesize that memory function improves system dynamics in communication networks. We predict that forgetting improves, rather than impedes, performance in situations where crowds use finite common resources. This is also a core problem in multi-agent learning: one agent's behavior may impact the outcome for another agent. The collective behavior of a learning system and individual reward can vary wildly and unpredictably. Human-controlled road networks exhibit similar traffic jams, though rarely with the same catastrophic consequences the multi-agent simulations suggest. An emerging property of the human-based system is adaptation and damping. We will use a multi-agent system in which each agent is implemented as a model of (relevant) human behavior, namely learning from observations and information obtained by others, and estimating quantities (or categorizing) based on current knowledge.

In our scenario, individual models repeatedly choose roads in a road network to get them from *home* to *work*. The time taken to traverse a road is a function of the number of other agents on the road when the agent begins to traverse the road.

The first contribution of this work is to apply human-inspired instance-based learning (IBL, Gonzalez, Lerch, & Lebiere, 2003) algorithm to the problem of multi-agent learning on the road congestion problem. The IBL model treats each trip from home to work as several instances of road segments and weighs instances based on several factors including recency when estimating time to traverse a road. We found that the agents using IBL do much better than the agents using an alternative, category-based model and about the same as agents using a communication intensive averaging model.

Second, this paper examines the dynamics of mixed human-machine systems. We demonstrate that the overall system performance is improved by even a relatively small number of the IBL agents and that there were no negative effects on either type of agent. When there were only a few IBL agents in the system, they performed relatively better than the ternary agents, but when there were many IBL agents all agents performed about the same.

Finally, the third contribution of this work is to show how changes to the underlying system, in addition to changes due to learning, impact performance. One might expect that more numerical approaches will respond to change more quickly and effectively than a learner relying on experiences. However, we found no evidence of this. Instead, IBL agents reacted very quickly and appropriately to underlying change, far better than the ternary agents. From these experiments we can see potential for using IBL in multi-agent learning settings and exploring these other settings is a key area of future focus.

Cognitive models have been combined to explain learning in team settings, primarily in a qualitative way (Sun, 2008). Reitter and Lebiere (2012) used decay in a model implemented within a cognitive architecture to show that decayed memory improves agent performance in a foraging scenario with multiple, communicating agents. Instance-based learning within cognitive models has been shown to explain human behavior in a number of cognitive decision-making tasks (Lebiere, Wallach, & West, 2000; Erev et al., 2010).

Scenario Framework

The framework for the scenario (Scerri, to appear) consists of agents A, places P and edges G over some number of iterations. Each agent $a \in A$ has some place, $p_{home} \in P$ where it starts each iteration and some place $p_{work} \in P$ where it must get to each iteration or day. To get to p_{work} it must use edges connecting places. Individual edges $g \in G$ connect exactly two places. The agents task is to get from p_{home} to p_{work} most quickly each iteration.

The time that it will take an agent to traverse an edge depends purely on the number of agents already on the edge when it gets to the edge. Specifically, we choose a simplified function to model limited resources that are affected by congestion: the time taken by an agent is $10 + n^3_{already}$, where $n_{already}$ is the number of agents on the edge when the agent reaches it. The simulation randomizes the order the agents execute so that in one iteration an agent might be the first on the edge and have a very short travel time and another iteration it might be tenth onto the edge and have a very long travel time, even if none of the agents change their routes.

This framework has two important features. First, the agents will get very different perspectives on speed of a edge, based on exactly when they get onto the edge. Hence, either many iterations or cooperation is needed to create an accurate model. Second, busy edges heavily penalize the agents, just a few extra agents on a edge will dramatically slow the last few agents down again making cooperation important.

For experimental purposes, there are only ten different p_{home} and p_{work} for 200 agents. This makes for more interesting traffic congestion problems, with more extreme cases, and requires more coordination among the agents, but, as was shown in (Scerri, to appear) does not qualitatively change the system dynamics.

In every iteration, each agent uses a model of the graph to plan a path from p_{home} to p_{work}. The agents use a standard A* algorithm (Russell, Norvig, & Artificial Intelligence, 1995) to do the planning based on their current model of edge traversal times. Agents are risk neutral, trying to minimize expected travel time. They then execute their plan without adapting to observed conditions. At the end of an iteration, the agents can communicate about what they observed. The model the agent plans with and the information it communicates are described below.

It is assumed that each agent plans selfishly, but communicates truthfully and cooperatively. We are interested in two primary metrics. First, the average time it takes for an agent to get from p_{home} to p_{work}. Second, as the agents build their models and adapt their plans to the changing models, the average transit time will change. As a secondary measure, we are interested in the change in average transit time over time.

Communication Network The agents are organized into a social network where they can only communicate directly with a small subset of the rest of the agents. Information is propagated through the network in a peer-to-peer manner. Unless otherwise noted below, we use a random network with degree 5 to connect the agents.

Model Path Planning

The cognitive agents have to choose a path that will most quickly get them to their destination, based on experiences so far and from experiences communicated from other agents. The optimal strategy might be one that considers likely plans by others and the changes they will make, given their previous experiences. However, this is typically infeasible, and theoretically the game-theoretic Traveller's Dilemma (Basu, 1994) applies: If one agent A anticipates another agent B's reaction, A would also anticipate B's anticipation of A's reaction, and so on. Rational players will end up with a poor solution (finite game), or they will be faced with a computation that does not scale.

Any accurate model of crowd cooperation needs to deal with limited communication bandwidth, learn (quickly) to achieve acceptable performance, adapt to changing network dynamics. In the following we describe two cognitively plausible models for reasoning about the road network, as well as an optimal one with high bandwidth needs. Earlier work has shown path-planning in a model in the cognitive architecture ACT-R (Reitter & Lebiere, 2010). However, here we focus on the memory components only and keep the path planning algorithm (A* planning) constant to facilitate meaningful comparison.

Categorizing Model: Ternary We include two cognitively plausible models at the agent level. The first, *ternary*, forms its belief about a road segment as a category: *slow*, *medium* or *fast*. The model keeps, for each edge, a normalized frequency distribution of the observed categories, decayed over time. Specifically, for each edge e, the model is $M_e = \{p_{slow}, p_{medium}, p_{fast}\}, p_{slow} + p_{medium} + p_{fast} = 1$. When an individual gets an observation of a particular category it adds β_{local} for a local observation and β for a communicated observation to the relevant p and then normalizes.

The models assume the most probable category, $\max M$ for planning. In the following experiments, an edge in a particular category is assumed to take time 300, 156 and 12 for p_{slow}, p_{medium} and p_{fast}, respectively, cor-

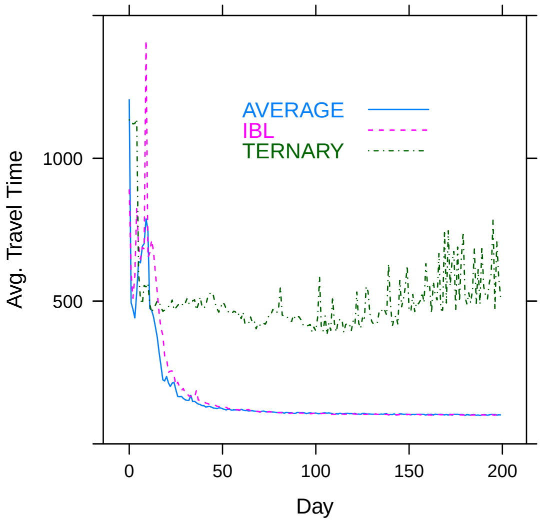

Figure 1: Ternary, IBL and Average models.

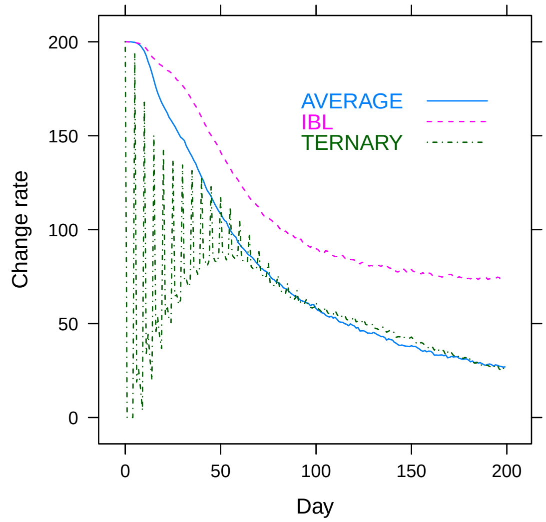

Figure 2: Rate of belief changes.

responding to the average time when approximately 3, 7 and 11 agents also use the edge reasonable approximation of the typical expectations. When $\max M$ changes for an edge, i.e., when the individual's belief about an edge changes categories, it communicates the new category to its direct neighbors in the social network.

Instance-based learning model The second model implements a cognitively motivated aggregation mechanism that forms their beliefs. As in the ternary model, its communications are quantized and occur whenever its belief about a road changes. The same A* algorithm is used to plan paths. However, this model's estimates about the speed of each road are based on instance-based learning. IBL stores a datapoint (episode) with the speed of a road whenever it is traveled or when agents receive a communication. (A commute involves many such roads.) A speed estimate can then be derived as the average of all episodes associated with the road, weighted by the episode's *activation*. Activation is determined by a function that rewards experience (a large set of episodes), but discounts older information (decay). Activation has been shown to predict the availability of information in human memory (Anderson & Lebiere, 1998).

In detail, activation of an episode e consisting of a road speed (utility) and time, $< u_e, t_e >$ is given as

$$A_e = (t - t_e)^{-0.5}$$

t is the current time. The decay exponent is the default that is empirically realistic in human experiments. Our implementation uses an highly precise approximation of the above activation function that omits to store all but the n latest episodes. If a road is represented by a series

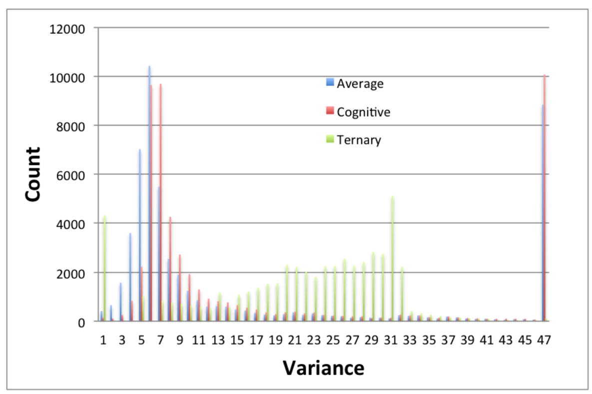

Figure 3: A histogram of the variance in estimates per road for each of the model types.

of episodes R involving the road, then the expected speed of a road, $U(R)$ is derived as

$$U(R) = \frac{\sum_{e \in R} u_e e^{A_e/T}}{\sum e^{A_e/T}}$$

$T = 0.25$ is a parameter (*temperature*). If R is empty, we assume a default speed, U_β for the road. The agent's performance is sensitive to U_β, which represents a measure of pessimism (we do not optimize U_β and choose 0.0 as the most optimistic value).

Instance-based learning and the activation function have several desirable properties in our context. Activation increases during early iterations and allows the model to quickly differentiate between fast and slow roads. Activation is less affected by presentation of changes concerning frequently travelled roads.

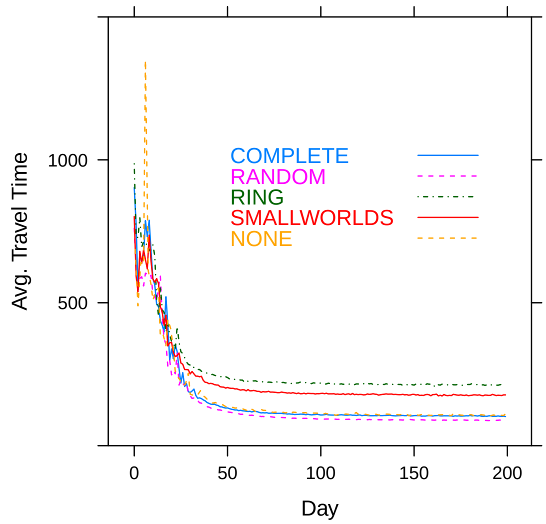

Figure 4: Average travel times for IBL model agents as the communication network is varied.

Averaging Model The Averaging Model is included to provide a form of non-cognitive empirical ceiling: it is information-hungry, assuming that communication is free and unconstrained. It is the simplest model an agent can have of the graph is to store the average time taken by agents traversing that edge. Since the utilization of an edge will change over time, a moving average is used to keep the model updated with respect to the current situation.

The agents estimate for an edge is simply $e'_i = \alpha e_i + (1 - \alpha)obs$, where e_i is the current estimate for the edge and obs is the new observation for the edge, whether communicated or observed locally. In this paper, we use $\alpha = 0.95$.

Empirical Evaluation

In this section, we empirically examine the three models on the congestion problem described in Section 2. The evaluation is split out into three parts, with each part aimed at looking in depth at one of the hypotheses introduced in Section 1. Unless otherwise stated, for each experiment below we use the following experimental parameters.

Instance-Based Multi-Agent Learning

The key challenge for multi-agent learning is that all the agents are simultaneously learning, making the learning environment non-stationary. Learning from instances in a non-stationary environment is not an intuitively effective technique. However, humans, who arguably use a type of IBL, are highly capable of learning in non-stationary environments. Our first experiments are aimed at looking at the performance of IBL on the congestion problem. Figure 1 compares the IBL, Ternary and Average models. Each model shows some improvement over time and some initial poor performance as the space is explored. The highly communication intensive and, for a human, computationally challenging Average model and the IBL model achieve about the same final level of performance and have about the same initially poor performance. Both do better than the Ternary model in the long run, although the Ternary model more quickly finds decent solutions.

Since the IBL and Average models end with about the same performance, it is tempting to conclude that they work in about the same way. However, Figures 2 and 3 show that they actually achieve the result with quite different dynamics. Figure 2 shows the average number of agents that change the path they take from the day before. The ternary model oscillates because beliefs take some time to change. More interestingly, IBL consistently changes more than Average. IBL agents change paths substantially more often, but the net result is the same as the Average agents. It is infeasible to determine exactly what is occurring, but it appears that IBL agents switch between approximately equal paths due to the noise in their relatively sparse data, while the Average agents have aggregated more data leading to more stable choices.

Figure 3 shows a snapshot of the variance in beliefs of the agents at the end of the 200 days. Specifically, for each road segment we computed the variance in the time the agents estimate it would take to traverse that road. These variances were then discretized and presented as a histogram, with variances > 50 put in the last bin for clarity. The higher the variance the more the agents disagreed about how long it would take to traverse the road. Each of the three models lead to distinctly different patterns. The Ternary case often has all agents in agreement and never has large disagreements between agents due to the way beliefs cascade across the network and because the agents only allow a road estimate to have one of three values. The Average model shows slightly lower variance overall than IBL, though the IBL has many more roads with very high variance, indicating complete disagreement. It is insightful to see that better performance was had when the agents had different models of the environment, many of which must actually be wrong. We can conclude that Average and IBL achieve approximately the same results, with very different algorithms and with distinctly different internal dynamics.

Conceptually, the cognitive model (IBL) does several things differently to Ternary. To try to understand what the cause of the different behavior was, we manipulated Ternary in several different ways. First, we artificially prevent Ternary agents from changing each step to mimic the IBL's preference for reusing previous paths. Second, we decay the learning rate so later data has less effect on Ternary, to mimic the way IBL instances aggre-

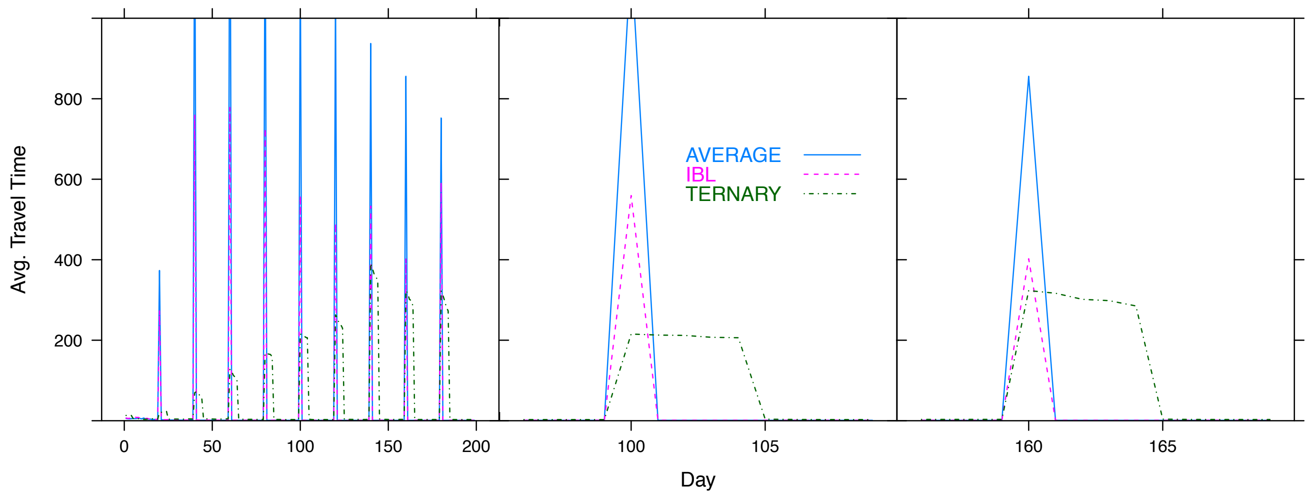

Figure 5: Travel times for different arrangements (lefts) of IBL and Ternary agents. Adding roads over time (right).

gate. Finally, we change the default value for Ternary for unknown roads to match the default for IBL. We found that each of these changes improved Ternary performance, but preventing them from changing each step had the biggest effect. The qualitative equivalent of this in human decision-making would be status-quo effects or confirmation biases, while IBL's implementation more directly reflects properties of human memory.

Figure 4 shows how communication networks influence the IBL agents. Curiously, blocking communication works similarly well as communication on fully connected and random network structures. These networks share information most evenly across the team, while ring and, to a lesser degree, small worlds networks compartmentalize information into neighborhoods. Although the effect is not very big, the data represents many simulation runs so the differences are not due to noise.

We see that complete, random networks do very well. A post-hoc explanation is that these networks enable the agents to communicate freely; agents have up-to-date information about congested roads. (The random networks were dense - each node has a degree of 5.) The networks without connections also do well, perhaps surprisingly so. Here, agents may adapt more slowly, and only to first-hand experience. In conjunction with the instance-based learner, this may also be a working strategy to avoid congestion. However, communication helps avoid a consistent initial spike, which we expect to be due to decision-making based on shared ground truth: everyone decides to use the fastest roads.

IBL and Ternary Models Interacting

IBL agents can be thought of as a simple model for how human learning might occur and Ternary agents can be thought of as a reasonable, low communication agent approach to cooperative learning. Future systems are likely to have humans and agents learning together and influencing each other. Hence, it is informative to look at what happens when IBL and Ternary agents are learning on the network at the same time. Varying the ratio of IBL to Ternary agents, we found that it takes relatively few IBL agents to give the whole system an improvement in performance. Having different types of learners in the same system not only does not hurt performance, it actually helps the weaker learners do better.

In the case of mixed Random graph networks of IBL and Ternary agents, we find that when there are only a few IBL agents they have a noticeable advantage over the Ternary agents, i.e., although they are using the same roads and are all interfering with each other, the IBL agents do relatively better. This advantage has disappeared when there are equal numbers of IBL and Ternary agents. The effect disappears smoothly as the number of IBL agents increases. If we think of IBL agents as being similar to humans and Ternary as being more like agents, this experiment hints that a small number of humans in an otherwise agent-dominated environment may do relatively better than the agents, and that they, as shown above, may improve the whole system's performance.

Disruptions

Intuitively, learning from instances is likely to behave differently to learning moving numerical estimates when there are changes to the underlying system. Here we look at two different types of disruption to the underlying system, the addition of roads and the addition of agents, and the effects on the dynamics for each of the agent types. In the first case, one new road is randomly added every 20 days. The resulting dynamics are shown in Figure 5 over 200 days (left) and just for an early (center) and a late (right) road addition. Both Average and IBL spike dramatically as they try to exploit the new road, but then go back to their original paths after finding it to be unhelpful – for most of them because they

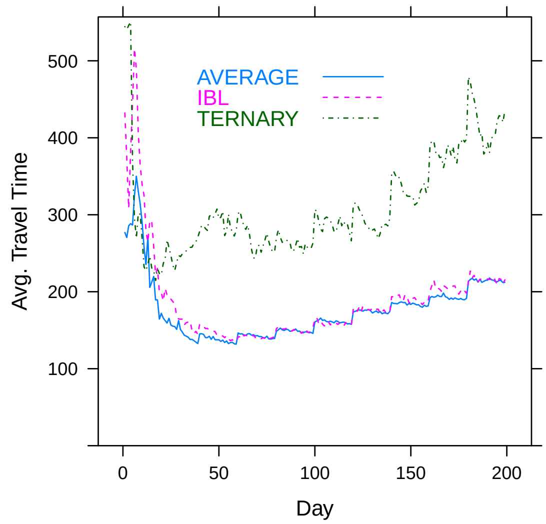

Figure 6: The impact of adding agents over time on average performance.

all tried it at once. The Ternary model is more robust because of the information sharing, but also takes longer to recover. Figure 6 shows the travel times as five new Ternary agents are added each 20 days, starting with 150 agents to make the result more comparable to other results. Both the Average and IBL agents jump when the new agents are added, but then smoothly improve performance. The Ternary agents are more dramatically effected by the change and do not adapt quickly. As the environment gets more congested and the original agents have built up more learning data, it is appears that IBL is more affected by the disruptions. This is unsurprising as its learning rate is effectively lower at this point.

Discussion and Future Work

The cognitive, IBL agents benefit from a relatively simple learning model, combining a preference for well-known roads and exploration of unseen roads. These cognitive agents can, with relatively limited communication volume, spread across the road network and efficiently use shared resources. What may be key to the cognitive agent's performance is limited sharing of knowledge: because agents do not have access to precise road utility estimates of their neighbors, and because they only receive updates when the neighbor's (quantized) beliefs change, they may arrive at heterogeneous conclusions about which roads are best. This leads them to spread out more, without sacrificing much individual performance. Under this scenario, agents do not need to misrepresent their knowledge states to their neighbors.

When combined in the same system, IBL agents and ternary agents actually helped each other rather than hurting performance. This is promising for future human-agent systems that will learn with distinctly different approaches. Notice that the agents are generally moving to a Nash Equilibrium, where, at least according to their local models, they have no incentive to change behavior. However even if the agents do reach an equilibrium, the outcome may be far from the socially optimal solution (Hagstrom & Abrams, 2001).

Understanding the emerging effects of cognitive decision-making in these networked simulations will allow us to spell out clear predictions to investigate crowd behavior empirically. The performance of cognitive agents that are based on empirically informed constraints of memory decay, instance-based learning and blending suggests that the mechanisms are not merely a rational adaptation to static information in the environment, but to dynamic resources and a social communication system. They enable us to maintain external, distributed memory without the devastating effects of cyclic, mutual adaptation.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum. Paperback.

Basu, K. (1994). The traveler's dilemma: Paradoxes of rationality in game theory. *The American Economic Review*, *84*(2), pp. 391-395.

Burgard, W., Moors, M., Fox, D., Simmons, R., & Thrun, S. (2000). Collaborative multi-robot exploration. In *IEEE Conference on Robotics and Automation, ICRA'00* (Vol. 1, pp. 476–481).

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., et al. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, *23*(1), 15-47.

Gonzalez, C., Lerch, F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591-635.

Hagstrom, J., & Abrams, R. (2001). Characterizing braess's paradox for traffic networks. In *Proc. IEEE Intelligent transportation systems* (pp. 836–841).

Kalyanakrishnan, S., Liu, Y., & Stone, P. (2007). Half field offense in robocup soccer: A multiagent reinforcement learning case study. *RoboCup 2006: Robot Soccer World Cup X*, 72–85.

Lebiere, C., Wallach, D., & West, R. L. (2000). A memory-based account of the prisoner's dilemma and other 2x2 games. In *Proceedings of the International Conference on Cognitive Modeling* (p. 185-193).

Reitter, D., & Lebiere, C. (2010). A cognitive model of spatial path planning. *Computational and Mathematical Organization Theory*, *16*(3), 220-245.

Reitter, D., & Lebiere, C. (2012). Social cognition: Memory decay and adaptive information filtering for robust information maintenance. In *Twenty-sixth AAAI Conference on Artificial Intelligence (AAAI-12).*

Russell, S., Norvig, P., & Artificial Intelligence, A. (1995). A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs.*

Scerri, P. (to appear). Modulating communication to improve multi agent learning convergence. In *Dynamics of information systems: Algorithmic approaches.*

Sun, R. (Ed.). (2008). *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation* (1st ed.). Cambridge University Press. Paperback.

Tesfatsion, L., & Judd, K. (2006). *Handbook of computational economics: Agent-based computational economics* (Vol. 2). North Holland.