Designing Socially Intelligent Agents For The Ultimatum Game

Peter M. Todd and Bernhard Borges
Center for Adaptive Behavior and Cognition
Max Planck Institute for Psychological Research
Leopoldstrasse 24, 80802 Munich GERMANY
ptodd@mpipf-muenchen.mpg.de, borges@mpipf-muenchen.mpg.de

Abstract

To build socially intelligent artificial agents, we must decide how much--and what kind of--intelligence to endow them with. Vriend (1997) has recently questioned whether adding reasoning will help or hinder the behavior of simple learning agents in social games. We show that adding the *right* kind of reasoning *will* help agents act more intelligently by learning more quickly.

Traversing the Spectrum of Reasoning Ability

How much intelligence does it take to be socially intelligent? Some researchers have proposed that social intelligence involves the rather complex ability to model the intentions and goals of the agents with whom one interacts socially (Byrne & Whiten, 1997). Others have suggested that simpler, "fast and frugal" mental heuristics may serve us in the social realm (Gigerenzer, 1997). Between these two points--high intelligence in the first case, and limited cognition in the second--lies a large range of possibilities. If we are interested in actually building socially intelligent agents, it would be good to have more guidance on the question of where we should aim our efforts: how much "smarts" will we have to build into our agents?

Game theory gives us a way to look at this question with a bit more precision. Here, the extremes of possible intelligence are easier to specify and agree on: On the high end, a fully- introspective agent could completely analyze any strategic situation encountered and unfold the back-and-forth turn-taking into the future to discover the best possible course of action. On the low end, a simply reactive agent could be programmed to react appropriately when it encounters strategic situations, without thinking. Of course, the latter type of agent will not be able to react properly to environmental stimuli for which it has not been programmed, so it lacks the central hallmark of intelligence: the ability to adapt. In some artificial settings in which the environment is known to be static, this may be sufficient, but in most fluid social situations such simple agents would quickly flounder. So we should aim for more intelligence--but do we need to go all the way to the far extreme of complete rationality and full introspection?

Recently, Vriend (1997) has addressed this question by elaborating the spectrum of reasoning or intelligence that game-playing agents might have. Slightly above the no-reasoning reactive agent, Roth and Erev, hereafter RE, (1995) have described "low rationality" agents that can adapt their behavior to a wide variety of game settings through simple reinforcement learning. In this case, game players merely try different actions according to some probability, and when they get positive feedback from the environment (the other agents they are playing against), they increase the probability of performing the same action in the future (and decrease the probability after receiving negative payoffs). Such simple agents can actually learn the "optimum" strategy (the one yielded by full introspection) in many situations, but this is not guaranteed, and when it does happen it can require a long period of playing and reinforcement.

Simple reinforcement learning may suffice for building socially intelligent agents in some situations, especially those where a long initial training period can be provided before the agents must in real-time where payoffs matter, and where the payoff structure of the environment is likely to change only slowly if at all. But in many cases, we would like to have agents that can learn how to behave more rapidly, and adapt to the changing strategies of their fellow social agents rapidly as well. To accomplish this, we can add more reasoning abilities to the agents, so that they can learn not only from their own immediate behavior and feedback, but also from "deductions" they make about their behavior and feedback. Thus, instead of merely using actual reinforcement from the environment, they can also use "virtual" reinforcement that they construct cognitively. It stands to reason that adding such reasoning abilities to learning agents should make them more intelligent, and allow them to learn the proper strategic actions more quickly.

But Vriend (1997) has shown that this assumption can in fact be false--adding reasoning to learning agents can sometimes hinder learning, rather than helping it. He presents a case in which a logically plausible form of reasoning is added to reinforcement learning players in the ultimatum game. Instead of helping these agents to learn the theoretically-expected strategy faster, this reasoning is "fooled" by an information asymmetry in the game and ends up producing agents who play a distinctly suboptimal strategy. As Vriend (1997, p. 7) says, "Adding reasoning about unchosen actions, and hence virtual updating, to the basic reinforcement learning model brings you closer to the fully introspective game-theoretic approach.... But it leads to outcomes that are farther away from the game-theoretic prediction. Hence, reasoning does *not necessarily* [his emphasis] improve learning." So should we take this cautionary note as a warning against adding reasoning to our social agents, and just stick with the simple reinforcement learning approach instead as the least risky option?

In a word, no--the problem with adding reasoning abilities to learning agents is one of figuring out *what kind* of reasoning to add. As Vriend points out, some kinds of reasoning will hinder learning--but as we have found, *most* kinds of reasoning, at least in the particular case of the ultimatum game that we and Vriend have considered, will indeed help learning, and make our socially intelligent agents more intelligent, more rapidly. In the rest of this note, we describe our simulations to investigate this question, and mention the directions to explore from here.

Simulating Learning And Reasoning In The Ultimatum Game

We explore the interactions of learning and reasoning in the context of the ultimatum game, a simple strategic situation in which social intelligence can come into play. The ultimatum game has been studied extensively in the game-theory literature, among other reasons because the behavior of people playing this game is quite different from the theoretically expected behavior. The game is simple: one player, A, is given a pie with a certain number of slices, and she must decide how many slices to offer to the other player, B, and how many to keep for herself. Player B receives this ultimatum and either accepts it, in which case the pie is divided accordingly, or rejects it, in which case both players receive nothing. The two players only play this game with each other once. The unique subgame perfect equilibrium of this game is for player A to offer the minimal amount (one slice), and for B to always accept this offer. But human players often offer nearly half the pie, and do not accept offers much less than this (e.g., Güth, 1995).

RE (1995) show how simple reinforcement learning (RL) agents can learn strategies similar to those employed by humans in the ultimatum game. To see if reasoning can help learning in this context, Vriend (1997) adds a simple, logical reasoning rule to the RL agents acting as player A (the offerer). (He assumes that players B all play an "accept anything" rule--another simplifying assumption to which we return in the conclusions.) If agent A offered amount X of the pie-slices to agent B last time, and this offer was accepted, then A assumes this means that all offers above X would also have been accepted. Hence A should reinforce not only its use of amount X as an offer, but also its use of all amounts above X, because A wants to learn to make offers that will be accepted. When this hybrid learning/reasoning agent is constructed and allowed to make several offers and learn from them, its behavior, rather than approaching the game-theoretic prediction of making the minimal offer, tends towards making offers in the middle of the possible range for X. While this is also similar to what humans do in experimental ultimatum game situations, it is even further from the optimal solution than is the behavior of an agent that only uses RL alone (against "accept-anything" B-players). Thus Vriend concludes that adding some reasonable forms of reasoning can in fact make learning worse.

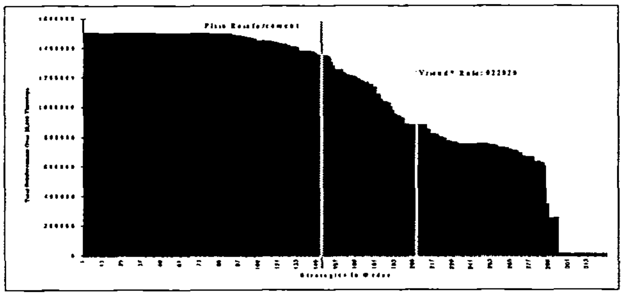
But how reasonable *was* the kind of reasoning that Vriend added? While it may have been logically correct and even rational, it was certainly not reasonable in an adaptive sense: any agent with that form of reasoning-plus-learning would, when placed in competition with a simpler RL-only agent, lose this competition, in the sense that the RL-only agent would learn to offer less, more quickly, and thus would keep more resources for itself throughout its game-playing life. But would other forms of reasoning, when added to the RL-only agent, give it a real adaptive advantage?

To find out, we tested a wide range of agents with different reasoning-plus-learning strategies. Each reasoning strategy, to be added onto the simple RL learner, was represented as follows: First, agent A can either choose its actions probabilistically from its learned set of reinforcements (coded as a 0), or she can always pick the action that has so far received the most reinforcement (1). Reinforcement can be calculated with a unit payoff (coded as 0), or with a fixed amount corresponding to what A will keep (1), or with a varying amount corresponding to what A *would* keep if she hypothetically offered different amounts in the virtual reinforcement cases (2). If agent A makes an offer X, and that offer is accepted, then A can either increase (2), decrease (1), or leave unchanged (0) the probability of offering X next time. Similarly, A can either increase, decrease, or leave unchanged the probabilities of offering all values *less* than X next time, and the same for all values *above* X. Finally, A can learn only when her offer is accepted (0), or else can also learn by negative reinforcement whenever her offer is rejected (1--but this case does not happen with the current B acceptance strategy). This gives 324 possible reasoning-plus-RL strategies to consider. A strategy similar to Vriend's (but slightly different, because he always chose all actions equally often during training) is coded by

022020 (reinforce the offer and all offers above it), while a plain reinforcement strategy (reinforce only the offer) is coded by 012000 or 022000.

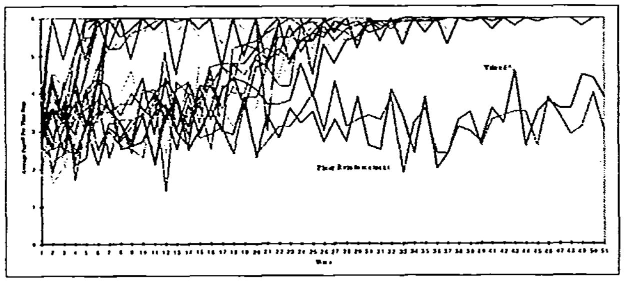
We ran all of these strategies 10 times, for 25,000 reinforcement periods each, with a pie-size of 6 slices, and looked first at what their total accumulated reinforcement was (which could maximally average 150,000). In Figure 1, we see the 364 strategies ordered by this total reinforcement, with the Vriend-like rule (Vriend*) and plain reinforcement both marked. We see that many strategies learned very quickly and so attained nearly the maximal amount of reinforcement possible. The Vriend*, however, faired rather poorly, coming in 213th place, and thus being beaten by approximately two-thirds of the whole set of possible strategies. Plain reinforcement did quite a bit better than this, coming in 150th. And many reasoning-plus-learning strategies did even better still.

Figure 1: Strategies Ordered By Total Reinforcement



If we look at the average step-by-step payoff in Figure 2 that the top strategies accumulate as they first learn in the initial 50 steps of training (i.e. in the first 50 ultimatum games they play), we can similarly see that most of these good strategies learn how to perform very quickly.

Figure 2: Strategies Ordered By Total Reinforcement



In contrast, the Vriend* and plain reinforcement start off slowly, hovering around the middle ranges of reinforcement. Eventually, plain reinforcement outstrips the Vriend*, as shown by Figure 1, but this is a rather slow process--adding the *right* kinds of reasoning to plain reinforcement can instead speed up the learning significantly.

Where To Go From Here

So with this simple competition we can conclude that while Vriend is correct in warning that *some* kinds of reasoning will not add to an agent's intelligence, we have given reason to believe that *most* kinds of reasoning *will* aid learning and adaptation. Thus, we should continue to search for the appropriate types of reasoning-plus-learning processes to use in different social settings confronted by our agents.

But how far are we still from the kind of highly-competent social intelligence that humans show? As mentioned earlier, RE (1995) simple RL agents can learn ultimatum game strategies similar to those employed by humans--in part because both the offerers *and* acceptors are learning how to play this game simultaneously, and so are adapting to each other over time. In contrast, Vriend's reasoning-plus-learning agents that end up also making offers like human players have got that way because the acceptors are fixed, and rather non- intelligent. Thus, Vriend has not really considered a social situation after all--he has just looked at adaptation to a fixed environment. We too have only described a fixed environment here, but in further studies we have considered the effects of co-evolving both offerer and acceptor strategies with our whole range of reasoning-plus-learning strategies, and there we have found once again that complex co-adapting processes can lead to agents with human-like ultimatum game behavior, but much more quickly than could reinforcement learning alone. So it may be that even when we aim towards modeling human-level social intelligence in simple games, adding the right kind of reasoning can get us there faster.

References

Whiten, A. and Byrne, R. W. eds. 1997. *Machiavellian Intelligence II*. Cambridge: Cambridge University Press.

Gigerenzer, G. 1997. *The Modularity of Social Intelligence*. In Whiten, A. and Byrne. R.W., 264-288.

Güth, W. 1995. On ultimatum bargaining experiments – A personal review. *Journal of Economic Behavior and Organization*, 27, 3, 329-344.

Roth, A. and Erev, I. 1995. Learning In Extensive-Form Games: Experimental Data and Simple Dynamic Models In the Intermediate Term. *Games and Economic Behavior*, 8, 164-212.

Vriend, N. 1997. Will Reasoning Improve Learning? *Economic Letters*. Forthcoming.