

What evolutionary game theory tells us about multiagent learning

Karl Tuyls^{a,*}, Simon Parsons^b

^a *Institute for Knowledge and Agent Technology, Maastricht University, The Netherlands*

^b *Department of Computer and Information Science, Brooklyn College, City University of New York, 2900 Bedford Avenue, Brooklyn, 11210 NY, USA*

Received 1 May 2006; received in revised form 8 January 2007; accepted 9 January 2007

Available online 26 January 2007

Abstract

This paper discusses *If multi-agent learning is the answer, what is the question?* [Y. Shoham, R. Powers, T. Grenager, If multi-agent learning is the answer, what is the question? Artificial Intelligence 171 (7) (2007) 365–377, this issue] from the perspective of evolutionary game theory. We briefly discuss the concepts of evolutionary game theory, and examine the main conclusions from [Y. Shoham, R. Powers, T. Grenager, If multi-agent learning is the answer, what is the question? Artificial Intelligence 171 (7) (2007) 365–377, this issue] with respect to some of our previous work. Overall we find much to agree with, concluding, however, that the central concerns of multiagent learning are rather narrow compared with the broad variety of work identified in [Y. Shoham, R. Powers, T. Grenager, If multi-agent learning is the answer, what is the question? Artificial Intelligence 171 (7) (2007) 365–377, this issue].

© 2007 Elsevier B.V. All rights reserved.

Keywords: Evolutionary game theory; Replicator dynamics; Multiagent learning

1. Introduction

In *If multi-agent learning is the answer, what is the question?* by Shoham, Powers and Grenager [20], the authors make a valiant effort to analyse the state of the field of multiagent learning, to summarise the results that have been achieved within the field, to discern the major research directions that have been followed, and to issue a call to arms. In short, Shoham et al. conclude that most work in multiagent learning can be placed into one of five “buckets” each of which is associated with a distinct research agenda (these descriptions are taken directly from the “caricatures” in Section 5 of [20]):

- (1) Computational: learning algorithms are a way to compute the properties of a game.
- (2) Descriptive: learning algorithms describe how natural agents learn in the context of other learners.
- (3) Normative: learning algorithms give a means to determine which sets of learning rules are in equilibrium with one another.
- (4) Prescriptive, cooperative: learning algorithms describe how agents should learn in order to achieve distributed control of dynamic systems.
- (5) Prescriptive, non-cooperative: learning algorithms describe how agents should act to obtain high rewards.

* Corresponding author.

E-mail addresses: k.tuyls@micc.unimaas.nl (K. Tuyls), parsons@sci.brooklyn.cuny.edu (S. Parsons).

In addition [20]:

Not all work in the field falls into one of these buckets. This means that either we need more buckets, or some work needs to be revisited or reconstructed so as to be well-grounded.

The authors also point out that research in multiagent learning is often unclear about which of these agendas it is pursuing—and that, in contrast one needs to be very clear about one’s aims—that the field cannot progress by defining “arbitrary learning strategies and analys(ing) whether the resulting dynamics converge in certain cases to a Nash equilibrium or some other solution concept”, and that the field needs to take evaluation more seriously, especially on some set of standard problems.

We basically agree with all of the points that we have quoted above, and below will expand on those that we feel our background leaves us best qualified to discuss. However, we do have one major point of disagreement with the views expressed in [20]. The point we disagree with is the idea that there is some overall *taxonomy* of research agendas into which all work can be slotted. This is not explicitly stated—all that Shoham et al. say is what we quoted above, that there are five distinct agendas to which some additional ones may need to be added—but the existence of an underlying taxonomy into which these additional categories can be slotted seems to be implied.

Our objection is neatly summarised by the following passage from McWhorter’s *The Power of Babel: A natural history of language* [15] in which the author tries to express how the original language, spoken by our common ancestors, became the many thousands of languages that descended from it. He starts by saying “I have implied that speech varieties have developed like a bush, starting from a single sprout and branching in all directions, each branch then developing subbranches, and so on . . .” before going on to explain that the inter-relationships between languages, the constant process of adoption of terms from one language into another, and the formation of dialects, creoles and intertwined languages means that [15, page 94]:

we might do just as well with another analogy, say stewing a good spring lamb stew without much juice (because the juice messes up the analogy). Clearly, one can distinguish the lamb from the peas from the potatoes from the carrots from the leeks from the rosemary leaves. Yet all of these ingredients, if it’s a good stew, are suffused with juice and flavor from the other items as well. Every now and then, you even encounter a piece of something that, covered with liquid and cooked out of its original shape and consistency, you have to work to figure out the original identity of . . . Overall, there is nothing in this stew that tastes or even looks like it would if you had just dumped it into a pot of boiling water by itself.

It seems to us that multiagent learning is such a stew and though it is very helpful to identify the various ingredients—especially if, to stretch the metaphor, some of them would be better taken out of the pot—to concentrate on the constituents misses some of the essence. It is the places in which the agendas that make up the stew meld into new things, things that cannot be put into a taxonomy because they are a mixture, that we often find the most interesting work.¹

That said, we should reiterate that we are largely in agreement with the agendas identified in [20], and in the next section amplify our agreement by examining all five agendas through the lens of evolutionary game theory (EGT), which is an area of multiagent learning in which we have been working, and one that is not much discussed in [20]. Following that exploration, we return to our point about the interplay between agendas, illustrating our discussion with some of our recent work.

2. The five research agendas from the perspective of evolutionary game theory

In this section we will use EGT to illustrate our reaction to the analysis of multiagent learning presented in [20]. To do this, we first consider what each of the five research agendas means in terms of EGT, taking them in the order that best fits our argument, rather than the order in which they are presented by Shoham et al.

¹ Though one has to be especially careful to be clear what one is doing at these junctures.

2.1. Normative and descriptive agendas

Our view of EGT is that it represents a move away from the normative agenda (in the terms of [20]) and towards the descriptive agenda. In particular, as summarised by [6], recent years have seen an important shift in game theory away from classical solution concepts such as Nash equilibrium, and towards EGT solution concepts such as the evolutionary stable strategy² (ESS) and the replicator equations. The obvious reasons for this are that the Nash equilibrium is hard to compute, often does not describe the best way to behave, both in terms of optimality and stability, and is not able to deal with highly dynamic situations. EGT [8,10,13,19] suffers less from these problems, and in our opinion, the descriptive approach that it embodies matches the overall goals of multiagent learning much better than the normative approach that is the preserve of traditional game theory (GT). In the subsequent sections we will explain the basis for these beliefs.

2.1.1. From game theory to evolutionary game theory

When John Maynard-Smith applied game theory to biology [13,14], and thus invented evolutionary game theory, he relaxed the premises behind GT. Classical GT is a normative theory, in the sense that it expects players or agents to be perfectly rational and behave accordingly [24,27,29]. In classical game theory, interactions between rational agents are modelled as games of two or more players that can choose from a set of strategies and the corresponding preferences. Game theory is thus the mathematical study of interactive decision making in the sense that the agents involved in the decisions take into account their own choices and those of others. Choices are determined by stable preferences concerning the outcomes of their possible decisions, and strategic interaction whereby agents take into account the relation between their own choices and the decisions of other agents.

Players in the classical setting have a perfect knowledge of the environment and the payoff tables, and try to maximise their individual payoff. However, under the biological circumstances considered by Maynard-Smith, it becomes impossible to judge what choices are the most rational. Instead of figuring out, a priori, how to optimise its actions, the question now facing a player becomes how to *learn* to optimise its behaviour and maximise its return, and it does this based on local knowledge and through a process of trial and error. This learning process matches the concept of evolution in biology, and forms the basis of EGT. In contrast to classical GT, then, EGT is a descriptive theory, describing this process of learning, and does not need the assumption of complete knowledge and perfect rationality. As a result, evolutionary game theory has steadily gained acceptance within economic game theory [6,7], and, given the close connection between game theory and much work in multiagent learning, we believe that it is an important foundation for the descriptive agenda. Evolutionary models describe how agents can make decisions in complex environments in which they interact with other agents. In other words, evolutionary models are expressly designed to work in exactly the kinds of environment that exist in the real world.

2.1.2. On the relation between the solution concepts from GT and EGT

To understand the importance of EGT for multiagent learning, it is helpful to summarise the relationship between solution concepts from GT and EGT. GT assumes that players will compute Nash equilibria and choose to play one such strategy. EGT assumes that players will gradually adjust their strategy over time in response to repeated observations of their own and others' payoffs. The *replicator dynamics* (RD) control this learning, specifying the frequency with which different pure strategies should be played depending on the mix of strategies played by the remainder of the population of agents playing the game. Strategies that gain above-average payoff become more likely to be played, and the RD models a process in which agents switch to strategies that appear to be more successful. Carrying out this analysis for a number of mixed strategies gives a picture of the whole game—at every point in the space of mixed strategies, it is possible to determine whether an agent should switch from its current strategy, and it is possible to plot a *direction field* that indicates the movement through the strategy space that such switches will lead to.

As an example, we consider the Prisoner's Dilemma (PD) game. In Fig. 2 we plot the direction field of the replicator equations applied to the PD. The direction field presented here consists of a grid of arrows tangential to the solution curves of the system. It is a graphical illustration of the vector field indicating the direction of the movement at every point of the grid in the state space. The x -axis represents the probability with which the first player will play defect

² A strategy is evolutionary stable if it is robust against evolutionary pressure from any strategy that appears. (In the context of a strategic game, "any strategy" is simply any of the other pure or mixed strategies that are available to an agent.)

		<i>B</i>	
		D	C
<i>A</i>	D	1	0
	C	5	3

Fig. 1. The strategic form of the Prisoner’s dilemma between the row player *A* and the column player *B*. *D* is Defect and *C* is Cooperate.

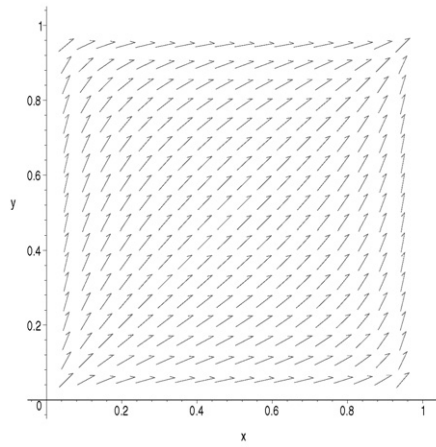


Fig. 2. Direction field plot of the Prisoner’s dilemma game.

and the *y*-axis represents the probability with which the second player will play defect. We could have made the same plot for the probability with which both players play their second strategy (cooperate).

The stationary points in the direction field for a given game are, clearly, the points to which groups of agents will evolve over time. Every Nash equilibrium of the game is such a stationary point, but the opposite does not hold. Thus the notion of a rest point for the RD on its own is not enough to give us an interpretation of Nash equilibrium in this context. As a result we introduce the criterion of *asymptotic stability*, which gives a kind of local test of dynamic robustness—local in the sense of minimal perturbations. The following is an intuitive definition of asymptotic stability: an equilibrium is asymptotically stable if the following two conditions hold:

- Any solution path of the RD that starts sufficiently close to the equilibrium remains arbitrarily close to it. This condition is called Liapunov stability.
- Any solution path that starts close enough to the equilibrium, converges to the equilibrium.

For a formal definition on asymptotic stability, we refer to [9].

Now, if an equilibrium of the RD is asymptotically stable—it is robust to local perturbations—then it is a Nash equilibrium [24]. This means that any solution that is asymptotically stable is also a Nash equilibrium, and so asymptotic stability is a weak refinement of the Nash concept. It is natural to ask whether ESS or asymptotic stability is also the strongest such refinement. Hofbauer, Sigmund and Schuster provided an answer to this question [10], proving that when a strategy *s* is ESS, then the population state it represents is stable in terms of the RD. This result gives us a way to refine the asymptotically stable rest points of the RD and it provides a way of selecting equilibria from the RD that show dynamic robustness, and are thus more convincing solutions than Nash equilibria (precisely because of that stability).

Turning back to our example in Fig. 2, we see that the Nash equilibrium and the ESS lie at coordinates (1, 1). All the movement goes toward this equilibrium.

		<i>B</i>	
		O	F
<i>A</i>	O	1	0
	F	0	2
		0	1

Fig. 3. Battle of the sexes.

Overall, then, we believe that in EGT, Nash equilibria continue to be relevant, indeed more relevant than Shoham et al. maybe credit. Since the equilibria of the RD are natural points at which groups of agents will end up, and we can imagine that the most interesting of these are the ones that are ESSs (since these are the ones at which agents will stay), and these points are Nash equilibria, what EGT tells us is that these Nash equilibria are a reasonable focus of attention. For a more detailed technical discussion we refer the reader to [22].

2.1.3. On to multiagent learning

Building models of agents that learn requires insight into the type and form of these agents' interactions with the environment and other agents in the system (including any humans). In much work on multiagent learning, this modelling is very similar to that used in a standard game theoretical model—players are assumed to have complete knowledge of the environment, are hyper-rational and optimise their individual payoff disregarding what this means for the utility and fairness of the entire population. In contrast, the basic properties of multiagent systems (MAS) seem to correspond well with those of EGT for two main reasons.

First of all, a MAS consists of actions and interactions between two or more independent agents, who each try to accomplish a certain, possibly cooperative or conflicting, goal. No agent will, in general, be completely informed about the other agents' intentions or goals, nor will it, in general, be completely informed about the complete state of the environment. EGT provides a mechanism in which the behaviour of such a system can be analysed. Second, a MAS is a dynamic system, in which agents will change their behaviour over time, in response to the behaviour of other agents. EGT offers us a solid basis to understand dynamic iterative situations in the context of strategic interactions, as shown above, and this fits well with the dynamic nature of a typical MAS.³

Not only do the fundamental assumptions of EGT and MAS seem to fit each other very well, but there is also a formal relationship between the replicator equations of EGT and reinforcement learning (RL) [2,21–23]. As an illustration of the latter, we provide an example of the relation between Q-learning, one of the most popular forms of RL, and EGT. More precisely, we present the battle of the sexes game in Fig. 3, together with the direction field and the Q-learning trace itself for two agents in Figs. 4 and 5. As we can see from the figures, the dynamics revealed by the trace are precisely those of the direction field. More technical details can be found in [22].

2.2. The computational agenda

The computational research agenda, as described by Shoham et al., is concerned with computing the properties of a game. As we have just seen, it is clear that evolutionary game theory can be considered to be a contribution to work on this agenda. What is less obvious is that we can use evolutionary game theory to compute other useful properties.

³ We might also argue that since a MAS is a reflection of the way that people behave in the real world—since agents will typically act on behalf of people—and as Gintis points out [8]:

Far from being the norm, people who are self-interested are in common parlance called sociopaths.

That it is odd to build systems of self-interested agents, and hence to use classical game theory to analyse MAS. However, one might also argue that by adequate construction of an agent's utility function, it can maximise utility but act in a way that is not obviously self-interested, and so game theory is a perfectly adequate tool for analysing MAS.

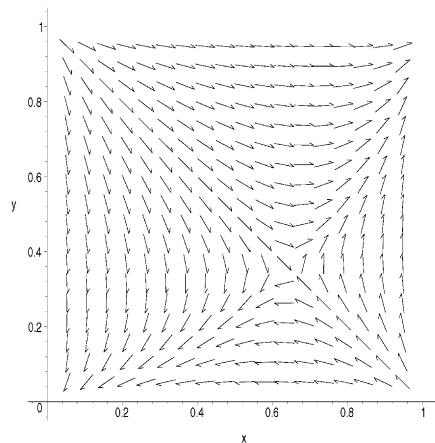


Fig. 4. Direction field plot of the battle of the sexes game.

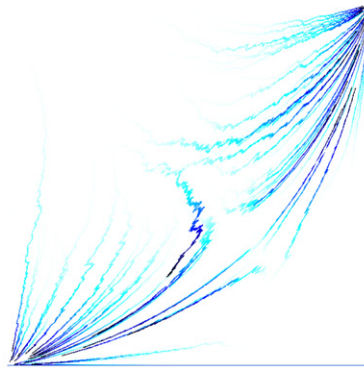


Fig. 5. Q-learning trace of the battle of the sexes game.

As an example of this, consider the work presented in [16], which used EGT to analyse the differences between two complex games played by the same set of players. The games analysed in [16] were two forms of double auction,⁴ the *continuous double auction* (CDA), in which matches between buyers and sellers can occur every time one makes an offer, and the *clearing house* (CH), in which buyers and sellers are matched only at some predetermined time point. The approach borrowed from the *heuristic strategy analysis* (HSA) of Walsh et al. [28] which assumes that in complex games, boundedly rational buyers and sellers will choose one of a handful of well-established bidding strategies rather than inventing some new approach. Given this assumption, it is then reasonable to model participants as choosing, using replicator dynamics, between some subset of this handful of strategies. In particular, [16] considered computing properties of the two double auction variants under the assumption that traders could choose between using the bidding strategy suggested by Preist and van Tol [17] and that proposed by Roth and Erev [4] which mimics human behaviour. (The point of the analysis was to be able to predict how human and automated traders would interact.) The results of the analysis are given in Figs. 6 and 7.⁵

The first set of results are, as suggested by Shoham et al., equilibrium results. As described above, one can interpret the stationary points in the replicator dynamics field as being the equilibria at which agents will end up given the initial mixture of strategies adopted by agents playing the game. These results, for example, suggest that in the CH if that, initially, half the agents are human—that is they use the Roth-Erev (RE) strategy—and half are automated—that is they

⁴ A double auction is one, like a stock market, in which there are multiple buyers and sellers. See [5] for more on double auctions.

⁵ Note, in passing, that these are true multi-player games. As [16] and [28] show, the results vary greatly as the number of agents change.

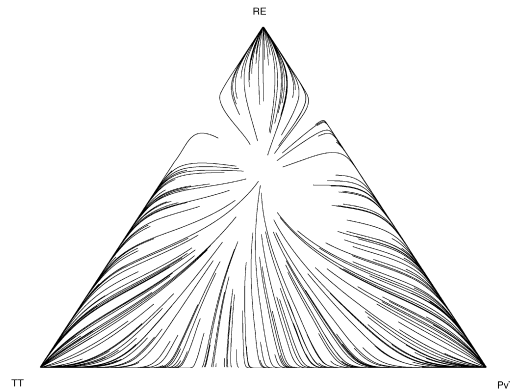


Fig. 6. Replicator dynamics direction field for CH with 10 agents.

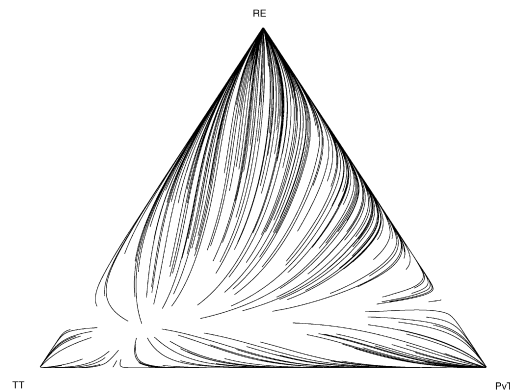


Fig. 7. Replicator dynamics direction field for CDA with 10 agents.

use the Preist and van Tol strategy (PVT)—then the system will reach equilibrium with all traders being automated. In contrast, in the CDA, the same initial conditions will lead to all agents using the RE strategy.⁶

However, as argued in [16], we can go further than this. If we assume that every starting point is, a priori, equally likely, then we can use the size of the “basin of attraction” for a given equilibrium—that is the set of points from which the replicator dynamics suggest agents will move to that equilibrium—as a measure of how likely a given equilibrium is. Combining that with the payoff that each strategy obtains at equilibrium, we can establish the expected payoff to agents in the game. [16] does this by assuming that all initial points (that is distribution of strategies between agents) is equally likely, but one could factor in any probability distribution over strategy distributions that one wanted to.

Of course the validity of the particular results obtained in [16] hinge on the underlying assumptions that are made by the heuristic strategy analysis, and so can, of course, be disputed. What seems beyond disputable, however, is that from the point of view of the computational perspective, EGT offers us a way to compute equilibrium behaviour and equilibrium properties of games, and we can use this information to provide comparisons between games.⁷ When working from a mechanism design perspective, in which one is interested in picking games to satisfy certain criteria, this kind of analysis can be very helpful.

To summarise our thoughts on the computational agenda, from the EGT perspective, this agenda is fundamental—EGT is primarily about computing the properties of games. Equally, every EGT analysis is a contribution to the computational agenda. Furthermore, this computation can both focus on both the identification of equilibria and the payoffs obtained by agents, and can be used to establish other properties of the interaction between agents.

⁶ The third pure strategy in Figs. 6 and 7 is “truth telling”, (TT), the strategy under which each trader honestly bids its true value for the good in question.

⁷ Though, as discussed in the next section, it is the equilibrium behaviour of a set of boundedly rational rather than perfectly rational agents.

2.3. *The prescriptive agenda(s)*

The final two agendas identified in [20] view multiagent learning as a mechanism for identifying how agents should behave in two different classes of problem—cases in which agents cooperate to provide the distributed control of some system, and thus share in overall team rewards, and cases in which agents compete against each other and look to achieve high personal payoffs. Again, we view them in a slightly different way to Shoham et al.

2.3.1. *Our view of the prescriptive agenda(s)*

Our view stems from our background as researchers who are interested in constructing multiagent systems. That is our concern is essentially an engineering concern—how can we construct systems that work? The kinds of system we are interested in are what one might consider orthodox multiagent systems (and our statement of what this means is, in essence, that which may be found in writings such as [30]). They are systems in which there are multiple threads of control—typically each thread of control is associated with some more or less independent entity which has some goals and the control of some resources that can be deployed to achieve those goals—and which are resource constrained—so that all the resources available to the various entities are not sufficient to achieve all of their goals. The dependencies between the resources and the goals mean that some of the entities will do best by cooperating, but the fact that the entities are independently controlled means that they are essentially in competition with one another.

From our perspective, then, the key issue is how to find the right way for these entities, represented by software agents, to act in order to both cooperate with the agents that it makes sense (in some well defined fashion) to cooperate, while doing their best (in a well-defined sense) to maximise their own position. This is why we consider the cooperative and the non-cooperative agendas under the same heading—in general an agent will be operating in a setting that is both cooperative and non-cooperative, and so work on just one agenda is only considering a part of the whole. (Though, of course, considering just a part of the whole is perfectly valid given the size of the task to understand how best to do this.)

We also take as paramount the need for the agents we are trying to engineer to be able to adapt.⁸ They need to adapt because their environments change over time, both inherently, because the sum total of all the agents in the system are not the only influences on their environment (no environment exists in a bubble), and because of the interactions they have with the agents in the system which, in general, will have a considerable effect on what it is best for them to do at any point in time.

In short, then, our focus is on how to obtain distributed control that adapts. For us this is what multiagent learning is all about, but, at the very least, it is one question that multiagent learning can help us answer. Furthermore, we believe that all of the research agendas identified by Shoham et al. contribute to finding answers to this question.

2.3.2. *Where we disagree with Shoham et al.*

We, considering ourselves to be completely orthodox in our views about multiagent systems, imagine that the systems we build will be composed of agents that are both cooperative, with respect to some of their peers, and non-cooperative, with respect to others of their peers. Furthermore, which peers a given agent cooperates with will, typically, change over time as the goals held by, and resources available to, the agent change over time.⁹ This is why we see that our overall goal is to establish mechanisms for adaptive distributed control, and why we believe that the two prescriptive agendas ultimately merge.

In contrast, Shoham et al. consider only the fourth of their five agendas—how to prescribe agent behaviour in cooperative domains—is concerned with adaptive distributed control, and so this form of control is most naturally modelled as a team game. Further, they suggest that [20]:

there is rarely a role for equilibrium analysis; the agents have no freedom to deviate from the prescribed algorithm.

⁸ Thus we agree completely with Shoham et al. that the interesting aspect is *adaptation* rather than *learning*, but also agree that arguing this point is a distraction from our main message.

⁹ The fact that agents are not wholly cooperative is often taken as the factor that distinguishes multiagent research from the narrower field of research in distributed artificial intelligence.

We disagree. Rather we believe that adaptive distributed control can arise perfectly naturally in non-cooperative domains (at least in domains that are non-cooperative in the sense we understand being used in [20]), domains that closely align with human society and domains in which agent behaviour is not prescribed.¹⁰

An example of this is provided by Rosenschein and Zlotkin, in their classic work on negotiation [18], with the example of the school run. Although every parent in a suburban neighbourhood derives a purely individual payoff from ensuring that their own children get to school and completing this delivery as quickly as possible—and so are not pre-programmed to be cooperative, nor subject to any team reward—these parents may still end up cooperating by reallocating the delivery of children amongst themselves in order to minimise their journeys. Even without joint rewards, even when considering only payoffs to individual agents, it may still be the case that solutions that involve some measure of cooperation, albeit cooperation that individually benefits every agent, will emerge. (This can, of course, be taken as the lesson from other overtly non-cooperative domains such as the Iterated Prisoner’s Dilemma [1].) Furthermore, from the right perspective, which might well be from sufficient distance that the naked self-interest of the agents is not visible, this self-organisation into some kind of recognisable pattern of delivery of children to school on time looks like a form of distributed control, one that adaptively manages the resources of parental time and car capacity. Finally, and now taking an EGT perspective, it seems clear to us that equilibrium analysis is important in examples like this. In EGT, establishing equilibria using the replicator dynamics tells us what states a given system will settle into over time and what intermediate states it will go through. In other words, equilibrium analysis is a tool for predicting how a system will adapt, not just what the equilibrium state will be.

There is one further thing to note, in passing. The example from Rosenschein and Zlotkin is what we might call a classical example, the kind of problem that was in vogue in the early days of multiagent systems. The designers of multiagent systems were presumed to be given a problem statement and an idea of what made a good allocation of resources, and expected to come up with strategies for agents to use that would enable agents with a range of properties to reach good solutions. This implicitly assumes that the system designers have access to the internals of agents, so that they can insert the necessary algorithms. A more modern view is of system design as *mechanism design*, in the usual economic meaning of that term [11]. In this kind of setting, designers are given the harder task of coming up with the rules for interactions between agents such that, no matter what algorithm the agents use to make their decisions, good solutions (where “good” is defined in terms of individual benefits, group benefit, or whatever one wants) will arise. This change, however, does not affect our argument.

3. Research agendas intermingle

As our discussion of the five research agendas has indicated, we believe that, from the perspective of EGT at least, it is not clear that these agendas are as distinct as suggested by Shoham et al. We can, for example, consider EGT to be both a means of computing properties of games and a mechanism for describing how agents will act in a game. Thus every application of EGT will be a contribution to several of the agendas, with the emphasis varying between applications. Thus, for example, the auction analysis we described above mainly contributes to the first and second agendas, while the load balancing work in [25,26] contributes to the first, fourth and fifth agendas. More generally, EGT can be thought of as being work on the computational agenda, work that focuses primarily on the payoffs to individual agents, and which is largely descriptive. Furthermore, since, in our view, all research in multiagent learning is, at bottom, concerned with how to build systems with adaptive, distributed control, the prescriptive agenda will often subsume all the others.

In a way, this position amplifies the importance of following Shoham et al.’s advice to researchers in multiagent learning about the need to be clear about exactly which agenda(s) they are contributing to. While in our view every piece of research in multiagent learning is, in some way, contributing to our understanding of how to create systems with adaptive distributed control, it is vital in this burgeoning field to be able to easily understand *how* each piece of research is making a contribution. The need for such clarity increases every time that we use EGT and find ourselves making contributions to several agendas at once. Indeed, we might be better of thinking of such work as being contributions to new agendas that arise from the intermingling of two or more of the five agendas from [20]. As such new agendas multiply, there are more possibilities to be picked through in order to make such an identification, and as

¹⁰ Noting that a staple of the intelligent agent literature is the notion that in the not-so-distant future, many agents will be used to handle mundane human tasks [3,12], and so agent societies may reasonably be assumed to mirror human societies.

agendas intermingle it becomes easier to mistake work on one mingled agenda from another, just as, to return to the stew analogy we borrowed at the start of this paper, it becomes harder to distinguish lamb from onion from carrot as the stew cooks longer.

4. Summary

In this paper we have taken an EGT perspective toward the five research agendas identified by Shoham et al. in [20]. In summary we can say we agree with the existence of the agendas, and the need for greater clarity in work on multiagent learning with respect to how it relates to these agendas, and thus how it contributes to the overall goals of the field. However, we feel that, as in our examples from evolutionary game theory, much work falls across the boundaries of several agendas, contributing to each in some way, and that there is nothing wrong with this. We believe that this “cross-agenda” work is only to be expected, and should be encouraged.

Looking at the individual agendas in more detail, we believe that work on the normative agenda has less to contribute to the overall goals of multiagent learning, which are ultimately concerned with computationally limited agents, than the descriptive agenda. It is the latter that will help us to build systems in which real agents and people interact with one another. Because these agents and people will be motivated at some level by their own self-interest, the payoff to individual agents has a more important role to play than equilibrium analysis, though certain classes of equilibria, such as the equilibria in replicator dynamics, are still informative.

Putting all of this together, our view is that the core activity of multiagent learning is in the area where Shoham et al.’s first, second and fifth agendas overlap. In this overlap we are concerned with computing the properties (agenda 1) of systems made up of real, rather than idealised, agents (agenda 2), whose decisions are motivated largely by their individual payoffs (agenda 5). It is because we believe that multiagent learning is an effective answer to the question of how to construct such systems that makes us interested in the field.

Acknowledgements

We would like to express our gratitude to our colleagues with whom we have cooperated over the years in work on multiagent learning. Furthermore, we wish to thank both the reviewers and the editors, Michael Wellman and Rakesh Vohra, for providing us valuable feedback on this document—we believe that this feedback has improved it considerably. The work described in this paper was partially funded by the National Science Foundation under grant NSF IIS-0329037.

References

- [1] R. Axelrod, *The Evolution of Cooperation*, Basic Books, New York, 1984.
- [2] T. Borgers, R. Sarin, Learning through reinforcement and replicator dynamics, *Journal of Economic Theory* 77 (1997).
- [3] A. Clark, *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*, Oxford University Press, Oxford, 2003.
- [4] I. Erev, A.E. Roth, Predicting how people play games with unique, mixed-strategy equilibria, *American Economic Review* 88 (1998) 848–881.
- [5] D. Friedlaender, The double auction institution: A survey, in: D. Friedman, J. Rust (Eds.), *The Double Auction Market: Institutions, Theories and Evidence*, in: Santa Fe Institute Studies in the Sciences of Complexity, Perseus Publishing, Cambridge, MA, 1993, pp. 3–25 (Chapter 1).
- [6] D. Friedman, Evolutionary economics goes mainstream: A review of “The Theory of Learning in Games”, *Journal of Evolutionary Economics* 8 (4) (December 1998) 423–432.
- [7] D. Fudenberg, D.K. Levine, *The Theory of Learning in Games*, MIT Press, Cambridge, MA, 1998.
- [8] H. Gintis, *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction*, Princeton University Press, 2001.
- [9] M.W. Hirsch, S. Smale, *Differential Equations, Dynamical Systems and Linear Algebra*, Academic Press, Inc, 1974.
- [10] J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, 1998.
- [11] M.O. Jackson, Mechanism theory, in: U. Devigis (Ed.), *Optimization and Operations Research*, in: *The Encyclopedia of Life Support Science*, EOLSS Publishers, Oxford, UK, 2003. The working paper version of this article includes a more comprehensive bibliography and some bibliographic notes.
- [12] P. Maes, Agents that reduce work and information overload, *Communications of the ACM* 37 (7) (July 1994) 30–40.
- [13] J. Maynard-Smith, *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, 1982.
- [14] J. Maynard-Smith, J. Price, The logic of animal conflict, *Nature* 146 (1973) 15–18.
- [15] J. McWhorter, *The Power of Babel: A Natural History of Language*, Harper-Collins, New York, 2003.
- [16] S. Phelps, S. Parsons, P. McBurney, Automated trading agents versus virtual humans: an evolutionary game-theoretic comparison of two double-auction market designs, in: *Proceedings of the 6th Workshop on Agent-Mediated Electronic Commerce*, New York, 2004.

- [17] C. Preist, M. van Tol, Adaptive agents in a persistent shout double auction, in: *Proceedings of the 1st International Conference on the Internet, Computing and Economics*, ACM Press, 1998, pp. 11–18.
- [18] J.S. Rosenschein, G. Zlotkin, *Rules of Encounter*, MIT Press, Cambridge, MA, 1994.
- [19] L. Samuelson, *Evolutionary Games and Equilibrium Selection*, MIT Press, Cambridge, MA, 1997.
- [20] Y. Shoham, R. Powers, T. Grenager, If multi-agent learning is the answer, what is the question? *Artificial Intelligence* 171 (7) (2007) 365–377, this issue.
- [21] P.J. 't Hoen, K. Tuyls, Engineering multi-agent reinforcement learning using evolutionary dynamics, in: *Proceedings of the 15th European Conference on Machine Learning*, 2004.
- [22] K. Tuyls, P.J. 't Hoen, B. Vanschoenwinkel, An evolutionary dynamical analysis of multi-agent learning in iterated games, *The Journal of Autonomous Agents and Multi-Agent Systems* 12 (2006) 115–153.
- [23] K. Tuyls, K. Verbeeck, T. Lenaerts, A selection-mutation model for Q-learning in multi-agent systems, in: *The Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, ACM Press, Melbourne, Australia, 2003.
- [24] F. Vega-Redondo, *Economics and the Theory of Games*, Cambridge University Press, Cambridge, 2003.
- [25] K. Verbeeck, *Coordinated exploration in multi-agent reinforcement learning*, PhD dissertation, Vrije Universiteit Brussel, Brussels, Belgium, 2004.
- [26] K. Verbeeck, A. Nowe, K. Tuyls, Coordinated exploration in multi-agent reinforcement learning: An application to load-balancing, in: *Proceedings of the Fifth Joint Conference on Autonomous Agents and Multi-Agent Systems*, Utrecht, The Netherlands, 2005.
- [27] J. von Neumann, O. Morgenstern, *Theory of Games and Economic Behaviour*, Princeton University Press, 1944.
- [28] W.E. Walsh, R. Das, G. Tesaro, J.O. Kephart, Analyzing complex strategic interactions in multi-agent systems, in: P. Gymtrasiwicz, S. Parsons (Eds.), *Proceedings of the 4th Workshop on Game Theoretic and Decision Theoretic Agents*, 2001.
- [29] J.W. Weibull, *Evolutionary Game Theory*, MIT Press, Cambridge, MA, 1996.
- [30] M. Wooldridge, *An Introduction to Multiagent Systems*, John Wiley & Sons, Ltd, Chichester, England, 2002.