Introducing Multiagent Systems to Undergraduates Through Games and Chocolate

Emma Bowring
University of the Pacific
Stockton
CA 95211, USA
ebowring@pacific.edu

Milind Tambe
University of Southern California
3737 Watt Way, PHE 410
Los Angeles, CA 90275, USA
tambe@usc.edu

ABSTRACT

The field of “intelligent agents and multiagent systems” is maturing; no longer is it a special topic to be introduced to graduate students after years of training in computer science and many introductory courses in artificial intelligence. Instead, the time is ripe to introduce agents and multiagents directly to undergraduate students, whether majoring in computer science or not. This chapter focuses on exactly this challenge, drawing on the co-authors’ experience of teaching several such undergraduate courses on agents and multiagents, over the last three years at two different universities. The chapter outlines three key issues that must be addressed. The first issue is facilitating students’ intuitive understanding of fundamental concepts of multiagent systems; we illustrate uses of science fiction materials and classroom games to not only provide students with the necessary intuitive understanding but with the excitement and motivation for studying multiagent systems. The second is in selecting the right material — either science-fiction material or games — for providing students the necessary motivation and intuition; we outline several criteria that have been useful in selecting such material. The third issue is in educating students about the fundamental philosophical, ethical and social issues surrounding agents and multiagent systems: we outline course materials and classroom activities that allow students to obtain this “big picture” futuristic vision of our science. We conclude with feedback received, lessons learned and impact on both the computer science students and non computer-science students.

1. INTRODUCTION

Since the first international conference on multiagent systems, ICMAS, held in 1995, to the international conference on agents and multiagent systems (AAMAS), 2009, the entire field of “intelligent agents and multiagent systems” as represented by AAMAS has matured significantly. In earlier years, students were introduced to this field as a special topic, only as a graduate course, after years of training in computer science, and after introductory courses in artificial intelligence. At the time, even the foundational principles of our field were unclear, and thus the only available syllabus was a set of advanced papers in multiagent systems.

Over the years, our field has gradually matured, and we are in a historic transition period. This is similar to fields such as robotics and software engineering that have matured to a point where undergraduates in computer science are able to take courses to develop relevant skill-sets in these fields. Agents and multiagent systems have similarly reached that critical mass. In fact, given the potential social impact of our field — society in general will need to interact with agents and multiagents on an increasing basis — introducing the fundamentals of our field to non-computer science (and non-engineering students in general) is also important.

This chapter focuses on this challenge, outlining three key issues that must be addressed. The first issue is facilitating students' intuitive understanding of fundamental concepts of multiagent systems. There are two obstacles: (i) while a number of textbooks have been written on agents and multiagent systems, there is not yet a standard set of concepts that are recognized as the foundation of the field and (ii) we need conceptual tools to help students understand multiagent systems without relying on an extensive computing background. In this chapter, we outline our approach to the lack of a standardized curriculum and more importantly, to address the issue of developing intuitive understanding of our science, we introduced science fiction materials and classroom games. These tools — science fiction and games — allow us provide students with excitement about multiagents, instilling a sense of wonder. The second issue is in selecting the right material, either science-fiction material or games: we outline several criteria that have been useful in selecting such material. In essence, we need to trade off conciseness of the material and its usefulness in the concepts taught. The key here is to ensure that a science fiction episode/story or game acts as "spice" to the main dish of the actual content from the field of agents and multiagent systems. Thus, these extra materials should not dominate our lectures, and should help rather than hinder teaching of the desired concepts. The third issue is in educating students about the philosophical and ethical issues that have surrounded agents and multiagent systems, as well as concerns about their future social impact. This is a crucial issue as it provides students with this bigger picture view of our field, educating them in the foundational debates such as the nature of intelligence. Science fiction is particularly important in allowing us to address this issue, providing a rich framework to construct exercises and discussions. For example, we outline a courtroom trial exercise based on a science-fiction TV episode, that allows students to debate the nature of intelligence, future rights of robots, and potential liabilities that agents designers may face in the future. Providing students with more creative and well-rounded courses that touched on philosophy, ethical and social concerns allowed us to attract a broader student audience that was keen on understanding the social context of computing.

The solutions outlined above have been practiced since 2006 via the coauthors' teaching multiple courses at two major universities in the United States to undergraduate students, both computer science majors and non-majors. For the computer science majors, we have taught three iterations of an upper-division course on multiagent systems. For non-majors, we have taught a variety of courses: including three "general education" courses that introduced concepts in multiagent systems and inspired students to take further computer science courses. We've also taught short seminars for "welcome week" where the courses were intended to encourage faculty-student interaction, and "parents weekend", where the goal was to allow parents to immerse themselves in an undergraduate course by attending short lectures taught by faculty. In our experience, all our student audiences face similar challenges in understanding basic concepts in agents and multiagent systems; however, the appropriate details that are needed to be covered for these audiences differ. In the case of computer science majors, the key is to provide sufficient details so they could actually implement the concepts as computer programs; for other audiences, the emphasis, particularly given the number of lectures and their format, may be to provide an understanding of the core concepts. Indeed, the key techniques introduced below appear to be useful to help teaching both computer science majors and non-majors.

2. MULTIAGENT SYSTEMS FOR UNDERGRADUATE STUDENTS

In introducing multiagent systems to undergraduate students, there are two main obstacles: (i) lack of a standard curriculum of fundamental agents concepts, and; (ii) lack of a common set of conceptual tools to provide the appropriate intuition and motivation for understanding key concepts. As far as a choice of appropriate textbooks, it is important to note that the existing set of textbooks on agents and multiagent systems provide a foundational contribution to our field[17, 13]. However, our courses provided a unique set of requirements. First, we needed to teach multiple audiences at different levels of preparation; we needed to introduce multiagent systems to undergraduates who were both computer science (CS) majors and non-majors and thus could not necessarily assume significant preparation in CS. Second, available textbooks did not necessarily cover key topics in multiagent systems that we wished to emphasize, such as teamwork, swarm behavior, distributed constraint optimization, behavioral game theory, coalition formation, agent-human interactions and others. Indeed, such differences in emphasis and materials was also revealed in an informal survey of syllabi offered in multiagent systems courses in key universities in the US and Europe --- we do not yet find conformity to one or the other textbook. Third, we wanted to include some cutting edge topics so students would recognize that not all questions in our field are settled. The goal was to allow students to understand that significant questions still remained unanswered and that this provided an exciting opportunity for further study in our field.

Our requirements led us to the conclusion that we could not rely on any of the available textbooks; yet providing undergraduate students detailed mathematical research papers to read was simply inappropriate. To address this we wrote a detailed set of class notes in intelligent agents and multiagent systems. The key was to write these notes for students of potentially diverse backgrounds, and hence to start from the basics. For example, students were not familiar with basic decision theory; hence the notes started out introducing decision theory and then built up to markov decision problems, and only then to partially observable markov decision problems (POMDPs) and finally to distributed POMDPs[10, 2]. The key here was to avoid getting into significant details of single-agent planning (or policy generation) algorithms[4, 6] and just provide fundamental concepts for single agents, and then get into the key multiagent algorithms. The end result is a course reader that is divided into three sections. The first section covers the fundamentals of agents and multiagent systems, e.g. BDI, decision theory and MDPs, game theory, auctions etc. A second section builds on the first, focusing more on multiagent collaboration and coordination. The third section is intended to focus on agent-human interactions. Appendix 1 provides an outline of our current syllabus.

A general lesson learned about the course reader from the feedback after the first iteration was the importance of including concrete examples of problems with solutions. In particular, in its first incarnation, our course reader did not include sufficient concrete examples. In later iterations, we provided a number of examples to illustrate key concepts as well as worked out examples at the end of each chapter. For instance, for sequential auctions, we provide a concrete example of a sequential auction and a worked out algorithmic solution for how agents would arrive at particular bids in such a sequential auction. Students have remarked that these examples have been particularly useful.

With respect to the common set of conceptual tools, the basic idea is to motivate the students, but more importantly, it is to help bridge the gap in students' understanding of multiagent systems. Consider concepts such as agent modeling, recursive agent modeling, the core in coalitional games, or risk aversion — these concepts are difficult to discuss in the abstract. Our goal was to find a hook into something that the students could intuitively understand, thus making it much easier to discuss these concepts with the students. Two separate types of tools have helped in this regard. First, science fiction stories, TV episodes and movies, have provided a social and fictional context for the discussion of basic elements of intelligent agent design. Many computer science students are already familiar with or fans of science fiction, and as such we build on this familiarity. For example, the notion of recursive agent modeling can be introduced in terms of science fiction episodes of a robot reasoning about a human's view of that robot (see below), which students find much easier to relate to than an abstract introduction to recursive agent modeling. Similarly, developing intuitions about the core in coalitional games or risk aversion is difficult in the abstract. Having students play games for something they value — we have found chocolate bars to be an ideal incentive — allows these intuitions to be developed. For example, discussions of risk aversion can be initiated by first having some volunteers decide whether they would prefer to take a chocolate or gamble for two chocolates and a penny. Some students prefer the certain chocolate, even though the expected payoff of taking the gamble is higher. This can help in discussions of decision theory and risk aversion. We discuss specific types of games in the following section and criteria for selecting them.

3. ISSUES IN USING SCIENCE FICTION AND GAMES

As discussed in previous section, we used science fiction and chocolate games to introduce key multiagent concepts; we now explain the criteria used in selecting such material and games.

3.0.1 Science Fiction Material Selection

Our selection criteria for the science fiction material included the following:

- It had to exhibit some topic of interest in the agents and multiagent systems arena, e.g. agent modeling, emotions, teamwork, agent interactions under uncertainty, etc. There had to be enough in-depth examination of this topic to enable students to build up an intuitive understanding of this topic, its relevance to multiagent systems, and some of the complexities that may arise in implementing it. In other words, the material needed to help us "bridge the gap" in understanding key concepts in multiagent systems.
- The story or film clip had to feature the robot or AI as an active participant in the plot or story.
- The story needed to present the robot/agent in a positive light. There is already a lot of popular science-fiction presenting a negative view of robots — combating this perspective with a positive view of robots/agents was considered important in order to motivate our students. (Unfortunately, meeting this requirement for all of our topics of interest proved to be quite difficult, and in one or two cases, we had to relax this requirement.)
- The story had to be short enough (maximum 30 pages); if a movie clip was to be used, it had to be short enough to be shown in 5-10 minutes at the beginning of class.

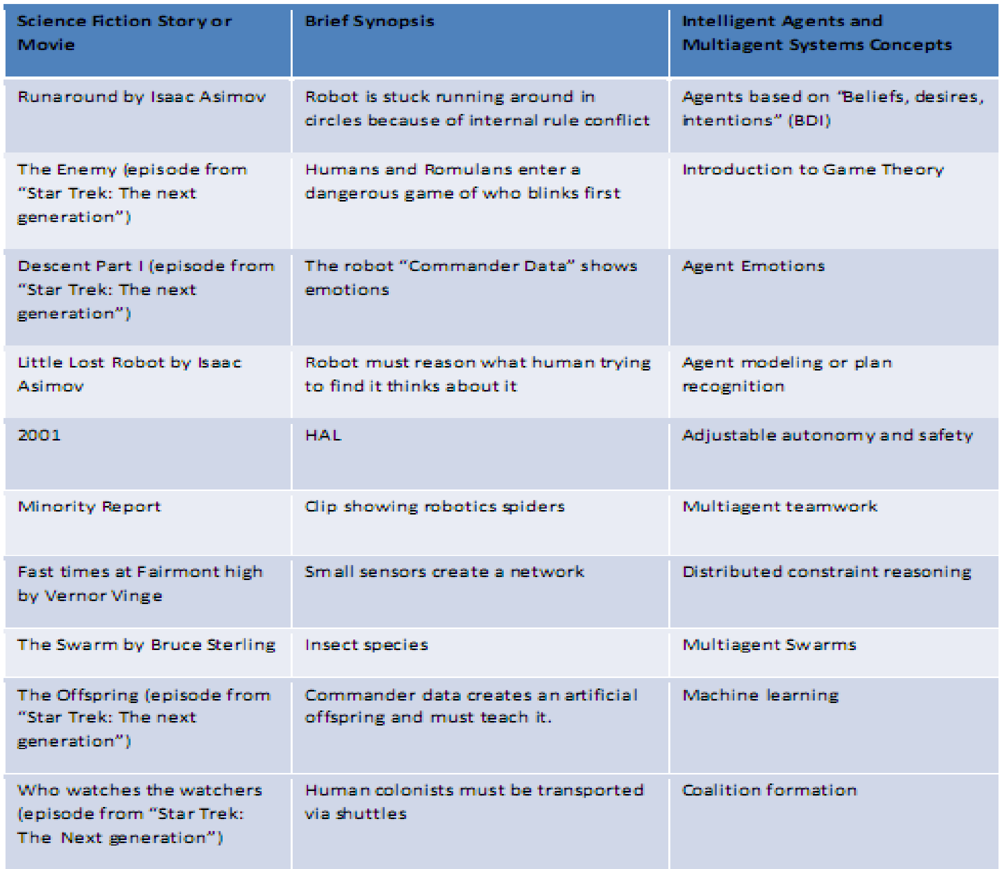
We chose several short stories by Isaac Asimov from his collection "Robot Vision" [1]. In addition to the short stories, the book also includes several essays by Asimov that provided useful reading material for the big-picture futuristic vision of our field as described later. We also chose episodes from "Star Trek: The Next Generation" that focused on the issue of agents, robots and intelligence.

Course lectures highlighted key aspects of agent behavior and functioning. For example, in the Asimov story, "Little Lost Robot" a robot (Nestor-10) intentionally tries to hide among a group of similar robots, while a human tries to run tests to isolate it from other robots. Nestor-10 is different from other robots purely in the rules it follows. In the story, Nestor-10 considers what the human believes the other robots will do in order to behave like all other robots — so it can blend in. Meanwhile, the human must devise tests that prevent Nestor-10 from blending in. In particular, she must infer the plan the robot is executing to pass the test; by observing the robot's actions, she can infer the plan the robot is executing, which in turn reveals whether the robot is Nestor-10 or not. Unfortunately, initially the human fails to recognize Nestor-10 because it anticipates the human's intentions, and foils those by continuing to successfully blend in. This story provides a fictional context for introducing basic concepts in agent modeling. In particular, the story provided three specific settings to investigate three particular aspects of agent modeling.

- The basic idea of a human trying to infer the plan a robot is executing by observing its actions provides an initial introduction to plan recognition.

- Nestor 10 must predict what other robots would do, and imitate their actions. This allows us to discuss how agents predict other agents' behaviors.
- Nestor 10 must recursively model what the human believes other (non-NESTOR-10) robots would do, enabling a discussion about recursive agent modeling.

Similarly, the Asimov story, "RUNAROUND" is based on a robot facing conflicting directives, which in effect brings up the notion of intention reconsideration and conflicting commitments. We can explain the behavior in terms of Belief-desire-intention (BDI) concepts[18, 7], and understand how to avoid such conflicts.



Science Fiction Story or Movie	Brief Synopsis	Intelligent Agents and Multiagent Systems Concepts
Runaround by Isaac Asimov	Robot is stuck running around in circles because of internal rule conflict	Agents based on "Beliefs, desires, intentions" (BDI)
The Enemy (episode from "Star Trek: The next generation")	Humans and Romulans enter a dangerous game of who blinks first	Introduction to Game Theory
Descent Part I (episode from "Star Trek: The next generation")	The robot "Commander Data" shows emotions	Agent Emotions
Little Lost Robot by Isaac Asimov	Robot must reason what human trying to find it thinks about it	Agent modeling or plan recognition
2001	HAL	Adjustable autonomy and safety
Minority Report	Clip showing robotics spiders	Multiagent teamwork
Fast times at Fairmont high by Vernor Vinge	Small sensors create a network	Distributed constraint reasoning
The Swarm by Bruce Sterling	Insect species	Multiagent Swarms
The Offspring (episode from "Star Trek: The next generation")	Commander data creates an artificial offspring and must teach it.	Machine learning
Who watches the watchers (episode from "Star Trek: The Next generation")	Human colonists must be transported via shuttles	Coalition formation

Table 1: Science Fiction Material used and concepts introduced

We outline in Table 1 some of the science fiction materials used and the concepts that were introduced using them. In all of these instances, students were either asked to read the story before class, or shown a short clip from the episode or film during class. In this way the story or film provided a context for a discussion about the key concept: i.e., why is agent modeling important, what are the difficulties in the problem, what are the key concepts offered in the filmic or textual "solution"? One major advantage of using science fiction in teaching multiagents was that it allowed us to introduce cutting edge topics and to instill a sense of wonder about our field.

To conclude, we established several criteria to select science fiction material for use in our classroom. We were often able to find science fiction material that met all our criteria; nonetheless, material that provided perfect fit for all our topics was not always available. The short movie clip of an angry robot from the movie "I, Robot" used in the lecture on emotions in agents does not touch on the usefulness of emotions or why it might arise so its use in "bridging the gap" is somewhat limited. The spiders from "Minority Report" do show teamwork and meet all our criteria except that they present a rather sinister view of robots. While we tried to meet all of our criteria with the science fiction we selected, we sometimes had to compromise when we couldn't find perfect material for a particular topic.

3.0.2 Choosing Games

Our criteria for designing games to play in the classroom included the following:

- The game had to force students to reason about something of value, but not cost too much to the instructors.
- The game had to exhibit some key characteristic of the topic studied in the lecture, just as in our selection of the science fiction episode.
- The game had to be short enough, lasting just 5-10 minutes at the beginning or in the middle of the lecture.

As mentioned earlier, having students play games for chocolates turned out to be an ideal solution for having objects of value that were not a significant cost to the instructors. Some of our lectures thus started out with a game at the beginning, and then focused on lessons from the game to introduce initial concepts. For example, lectures on decision theory and risk aversion started out with a game involving a gamble for chocolates. As mentioned earlier, this game involved first having some volunteers decide whether they would prefer a chocolate for sure or gamble for two chocolates and a penny. Playing out this game at the beginning of class led to a discussion how these choices were arrived at, and thus led into decision theory. Figure 2 illustrates a chart that is used to discuss students' preferences and compare their attitudes to risk.

In some variations of the course offerings, we also used chocolate games for introducing game theory. For example, we started off introducing prisoner's dilemma with the payoff function as shown in Figure 3. In particular, each pair of students was given a textual description of the game (not the actual game matrix) as shown below:

- Each player has a choice: Cooperate or Defect. Write down your choice as 'C' or 'D' on the piece of paper provided to you. Do not communicate with the other player.
- If both players cooperate, both players will get two chocolates each.
- If both defect, one chocolate each

- If one defects and one cooperates: the player who defects gets three chocolates, and the player who cooperates gets zero.

Risk Neutrality & Aversion

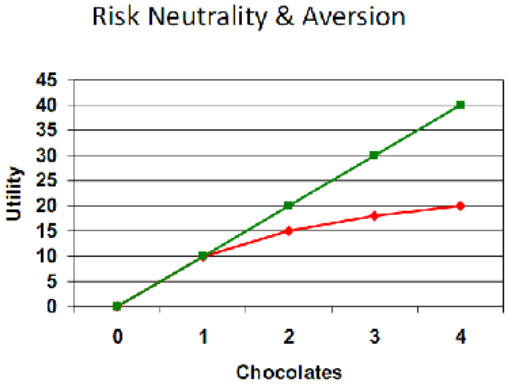


Figure 2: Results of a chocolate game to discuss risk neutrality and risk aversion; we discuss in class how the red line vs. the green line shows the change in utility with risk attitude.

The students were then asked to choose to cooperate or defect. The students' choices and their reasoning behind their choice then led into an introduction to game theory.

A key point to note is that in playing games, many different outcomes are possible. The diversity of outcomes often provided an important opening for deeper discussions and an introduction to more complex topics. For example, students may have different risk attitudes in the chocolate game used to introduce decision theory; this clearly opens up a discussion of differences in risk attitudes. Similarly, while the rational choice in the prisoner's dilemma game is for both players to defect, in the classroom, students may not necessarily play this ”rational strategy”. Indeed, in playing the prisoner's dilemma game, students quite often cooperate rather than defect. Such cooperation not only allows us to underline the paradox of prisoner's dilemma, but also provides an opening for further introduction of behavioral game theory.

In summary, we used chocolate games in introducing the following concepts:

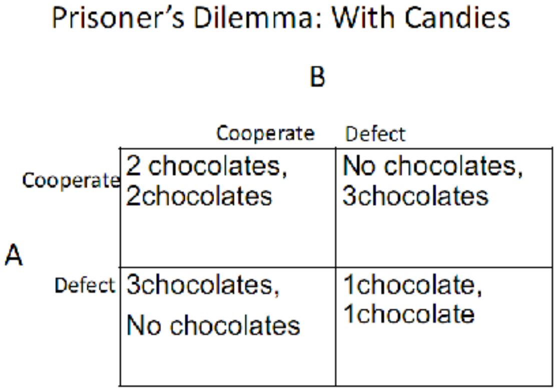
- Decision theory: Introduce expected value and risk averseness as discussed earlier.

- Game theory: Playing prisoner's dilemma using candies as discussed earlier.

- Auctions: Introduce Vickery auctions, sequential auctions using fake money and chocolates as items to be auctioned off. To ensure that the fake money had value, there was an exchange rate for the fake money for some small candies, so students would not always bet all their money on the chocolate auctioned. For example for introducing Vickery auction, four students were each given fake $100. They had to bid on a big chocolate bar. However, to ensure that they will not all bid $100, they were also told that each $16 bought them a small candy. In other words, if they paid $68, they could cash in their remaining $32 to buy two small candies. This led to more variations in students' bids.

- Coalition formation: To introduce concepts such as imputations and the core in coalitional games, we used chocolate games in class yet again; the students were given different characteristic functions in terms of chocolates and had to form coalitions. Multiple different groups of students were given different problems, e.g. in one case the core was empty, leading to very difficult negotiations.

Prisoner's Dilemma: With Candies



Prisoner's Dilemma: With Candies

		B	
		Cooperate	Defect
A	Cooperate	2 chocolates, 2chocolates	No chocolates, 3chocolates
	Defect	3chocolates, No chocolates	1chocolate, 1chocolate

Figure 3: Chocolate game to illustrate the prisoner's dilemma game.

One key principle used was to ensure that there were not too many different tools used in the same lecture. So for example, if a science fiction story or TV episode was used, then we did not simultaneously use a chocolate game. Using too many ideas could end up distracting students rather than helping them learn the concept of interest.

4. THE BIG PICTURE: PHILOSOPHICAL, ETHICAL AND SOCIAL ISSUES

The third issue that we need to address in teaching agents and multiagent systems to undergraduate students is to introduce these students to social, philosophical and ethical issues surrounding agents and multiagent systems, as a means of going beyond basic computer science aspects. Our goal was also to engage students in active learning[3] and teach them about the challenges of defining agents and multiagents as a field and the core issues of what it means to be intelligent. The aim was also to tie the material in class to a broader context and (hopefully) make connections to other classes/areas of interest, e.g. philosophy. There were two separate activities undertaken in this context, as discussed below.

4.0.3 Trial of a Robot

A series of two lectures focused on the trial of a robot to bring up fundamental philosophical arguments surrounding agents. Here we showed students half of the episode “The Measure of a Man” from “Star Trek: The Next Generation.” Commander Data is an android who is put on trial to determine whether or not he has rights. In particular, if the ruling of the trial is that Commander Data has no rights, then he would be immediately dismantled for further investigation — with a slim chance of being put back together.

The trial took place in two separate lectures. In the first lecture, students were shown the first half of the episode that chronicles the events leading up to Data’s trial. We then divided the class into four teams and gave the teams time to strategize/coordinate. Each team was assigned to argue either the pro or con of one of these two issues:

- Is commander Data intelligent? self-aware? sentient?
- Does Commander Data have rights? If Data creates art, who owns it? If Data kills someone, who is responsible?

We provided the students with a list of readings, supporting both pro and con positions. On one side were Turing’s “Computing Machinery and Intelligence” [16] and Asimov’s essays [1] and on the other were readings such as “Can a computer have a mind” [11] and Searle[14]. To prepare for the trial, students had to do the following homework:

As homework, you need to read 2 readings from the list provided below, at least one must be non-Asimov. Then, write a short list of three supporting arguments for the position you have been assigned. Your arguments can either be direct support or refutations of likely counter-arguments that the opposing team will make. For each of your points, please provide some support based on a reading. 1 point of extra credit can be earned on this assignment if you bring in an additional credible source to support your argument and provide a citation.

In the second lecture on this topic, students re-enacted the trial in class. During the trial, students had to speak up in support of their position based on their writeup submitted as part of their assignment. As instructors, we were not completely sure how this trial would work out. Our observation was that students were really passionate in support of their positions and had researched many extra sources. Several new and innovative arguments were brought to the floor, e.g. one student brought up the national historic preservation act to argue that Commander Data could not be dismantled because he was a historic engineering artifact.

During one iteration of this class, we had a third lecture where the author of this episode, Melinda Snodgrass, visited our class and gave an invited lecture. Starting with providing students the underlying motivation for this particular episode, the Dred Scott Supreme Court decision in the United States, the author went on to describe the role of robots in science fiction. She explained how she had investigated different human characteristics using robots, e.g. what would it mean for a robot to be leading a group of humans and what leadership entails for a robot. This contributed to both the philosophical fundamentals, but also the sense of awe about intelligent agents to inspire students to further continue studying in our field.

4.0.4 Futuristic Concerns about Agents

Science fiction writers, mass media reporters and some non-fiction writers have presented stories and scenarios that express concerns about future intelligent agents and multiagent systems and the harm they may potentially cause humans[1, 9]. For example, in some science fiction movies or stories, we see agents/robots developing their own agenda and ultimately even killing humans. In others, agents are specifically designed to intrude on people's privacy. In still others, agents may malfunction and create problems. In other words, science fiction writers have been worried that agents may cause us harm in any one of a number of ways: emotional, financial, physical, societal and so on.

The traditional view of AI education has been to ignore such concerns, in part because addressing such concerns is sometimes outside the scope of AI researchers' expertise and in part because some of these concerns appear too futuristic. Fortunately, AAAI (Association for Advancement of Artificial Intelligence) has now recognized this concern, and it has recently formed a panel to study and report on this issue. Furthermore, some researchers have already begun focusing on research on so called "Asimovian agents", explicitly inspired by Asimov's three laws of safety[5]. Our goal, however, was to bring these concerns to the notice of our students, and more importantly, to understand what mechanisms, engineered either at design time, or some via the legal system or societal convention would help address these concerns. To that end, students were again divided into teams. Each team was given an assignment to focus on a movie or a short story where agents caused harm. Specifically, the assignment asked students to first identify a design approach based on the techniques learned in class that could be used to build the agent. Next they were asked to:

Identify key scenes in the movie/story where the agent causes harm. Using your approach explain how it could give rise to such a harmful action, if not properly designed. Pick two scenes in particular and answer two questions: (i) explain what might cause the problem shown to arise. (ii) how realistic is the problem described?

The next step was crucial: identifying potential design decisions, social conventions, or laws that would significantly reduce or eliminate the potential for such harm. The key was to allow agents to exist functionally; in particular, very strong restrictions could completely eliminate agents' usefulness.

The results of this assignment were projects that outlined a potential design for agents in a variety of popular movies and stories and a range of potential solutions. In one case agents' emotions were held responsible, and thus providing agents with emotions was shown to be of concern. In another case, a student group analyzed Asimov's three laws, pointing out how these laws would need to be specialized before we delve further into the "Asimovian agents".

5. FEEDBACK AND LESSONS LEARNED

Results from student feedback on the courses have been encouraging. Students provided overwhelmingly positive responses, not only about the course material but also about the teaching technique of using science fiction. More specifically:

- While exact numeric evaluation scores from students on the course may be difficult to evaluate in the abstract (without having an appropriate baseline), it is useful to note that the scores received have been among the highest that the instructors have received among all courses they have taught.

- We provided students with additional questions to evaluate the role of science fiction: 16/20 students who provided feedback thought that the science fiction really added value to the material taught. Students commented that without the science fiction they would not have taken the course.

- At least some students (from our classes focused on non-majors) have suggested that they will change or have changed their majors to computer science; students from the "welcome seminar" have continued to correspond via email querying about agents and multiagent systems.

- Half a dozen students from the upper division computer science class have joined AI research labs to pursue research in agents and multiagent systems, either as undergraduate researchers, PhD students, or research programmers.

- A few of the students also immediately followed up the courses to take more in-depth courses in computer science.

Given the success of our courses, it is useful to step back and understand some of the key factors that in our view contributed to this success. These include:

- We were teaching material for which there was no clear textbook, bringing in cutting edge research material into our courses, and simultaneously using new teaching tools. This required significant amounts of planning, even much more so than traditional courses; we were fortunate enough to have been forced into this planning a year in advance partly by our faculty colleagues who engaged us in lively discussions.

- Figuring out the right science fiction material to use given our criteria proved to be extremely difficult. If the trend of using science fiction in AI were to be carried forward, it would help to build up a collaborative database of science fiction materials utilized.

- Finally, colleagues in AI and in particular agents and multiagent systems from around the world were extremely supportive of this effort, providing both encouragement and pointers to relevant materials.

6. SUMMARY AND RELATED COURSEWORK

It is now time to step up to the challenge of introducing agents to undergraduate students. To that end, we have used science fiction and games in the classroom as a means to bridge the gap in students' understanding of agents and multiagent systems, and to generate excitement about our science. Our mission was not only to teach students multiagent systems, but also to provide students with a broader view of the social and cultural context of the development of intelligent agents, including a discussion about the ethical, philosophical and societal issues relating to this work. To that end, since the fall of 2006, we have taught several courses at two different universities, both to majors and non-majors.

This chapter outlined the courses taught using this framework, provided an overview of our classroom teaching techniques in using science fiction and classroom games, and discussed some of the lectures in more detail as exemplars. We discussed the overwhelmingly positive student response and provided concrete examples of students turning to multiagents research as a result as well as to changing their majors to computer science.

There are other courses offered in other universities that use science fiction as a way of introducing science in general. Prof. Barry Luokkala of Carnegie Mellon University, a professor in the physics department, uses science fiction for introducing science. Other similarly offered courses include:

• NIH’s program called "Science in the Cinema"
http://science.education.nih.gov/cinema

• The American Chemical Society’s similar program:
http://www.scalacs.org/ScienceCinema/.

In particular, these programs also use film and world leading scientists commenting on the films to educate students and the general public. Bringing some of these approaches to teaching agents and multiagent systems has resulted in enthusiastic and strongly positive student response.

7. REFERENCES

[1] I. Asimov. Robot Vision. Penguin publishers, 1991.

[2] R. Becker, S. Zilberstein, V. Lesser, and C. V. Goldman.
Transition-independent decentralized Markov decision processes. In Proceedings of the Second International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-03), pages 41–48, 2003.

[3] C. Bonwell and J. Eison. Active learning— creating excitement in the classroom. ASHE-ERIC Higher Education Report, 1, 1991.

[4] A. Cassandra, M. Littman, and N. Zhang. Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes. In Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97), pages 54–61, 1997.

[5] D. Gordon. Asimovian adaptive agents. Journal of AI Research, 13:95, 2000.

[6] L. Kaelbling, M. Littman, and A. Cassandra. Planning and acting in partially observable stochastic domains. Artificial Intelligence, 101(2):99–134, 1998.

[7] H. J. Levesque, P. R. Cohen, and J. Nunes. On acting together. In Proceedings of the National Conference on Artificial Intelligence, pages 94–99. Menlo Park, Calif.: AAAI press, 1990.

[8] S. Marsella and J. Gratch. A step toward irrationality: Using emotion to change belief. In Proceedings of First International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS-02), 2002.

[9] B. McKibben. Enough: Staying human in an engineered age. Owl books, New York, 2003.

[10] R. Nair, D. Pynadath, M. Yokoo, M. Tambe, and S. Marsella Taming decentralized POMDPs: Towards efficient policy computation for multiagent settings. In Proceedings of the

Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03), pages 705–711, 2003.

[11] R. Penrose. The emporer's new mind: Concerning computers, minds and the laws of physics. Oxford University Press, 2002.

[12] N. Reilly. Believable Social and Emotional Agents. PhD dissertation CMU-CS-96-138, Carnegie Mellon University, Pittsburgh, PA, 1996.

[13] S. Russell and P. Norvig. Artificial Intelligence: A modern approach. Prentice Hall, 2003.

[14] J. Searle. Minds, brains and programs. Behavioral and Brain Sciences, 3:417–457, 1980.

[15] M. Tambe, E. Bowring, H. Jung, G. Kaminka, R. T. Maheswaran, J. Marecki, P. J. Modi, R. Nair, S. Okamoto, J. P. Pearce, P. Paruchuri, D. V. Pynadath, P. Scerri, N. Schurr, and P. Varakantham. Conflicts in teamwork: Hybrids to the rescue. In Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), July 2005, 2005.

[16] A. Turing. Computing machinery and intelligence. Mind, 49:433–460, 1950.

[17] M. Wooldridge. An Introduction to Multiagent Systems. John Wiley & Sons, 2002.

[18] M. Wooldridge, J. Muller, and M. Tambe, editors. Intelligent Agents, volume 2 of Lecture Notes in Artificial Intelligence 1037. Springer-Verlag, Heidelberg, Germany, 1996.

Appendix I: Detailed Syllabus

1. Lecture 1: Course intro, syllabus and course structure, what is an intelligent agent
   Homework: Read Asimov's short story "Runaround" for next lecture

Part I: Fundamentals of agents and multiagent systems

2. Lecture 2: Beliefs, desires, intentions (BDI), Satisficing and bounded rationality, Begin BDI logics

3. Lecture 3: BDI logics continued, BDI architectures (e.g. PRS, Soar), reactive plans

4. Lecture 4: Decision theory I; Making simple decisions under uncertainty; risk averseness, risk neutrality; begin Sequential decisions under uncertainty, Markov Decision Problems (MDPs)
   Use Asimov "Runaround"

5. Lecture 5: Decision theory II; MDP Value iteration; Introduction to Partially Observable Markov Decision Problems (POMDPs)
   Homework: Read Asimov's Chapter 1 (for history of science fiction)

6. Lecture 6: (INVITED LECTURE) "History of science fiction"

7. Lecture 7: Game theory I: Normal form and extensive form games, Prisoner's dilemma, Chicken, Dominance, iterative dominance, Nash equilibrium, Mixed strategy Nash equilibrium
   "Star Trek: The next generation" episode "The enemy" (Season 3)

8. Lecture 8: Game theory II: Iterative Prisonner's dilemma, Stackelberg games, Bayesian Games, Harsansyi transformation
   Need a new science fiction episode for mixed strategies

9. Lecture 9: Agents and emotions; Moral Emotions; Behavioral Game Theory Intro
   "I, Robot" clip

10. Lecture 10: Auctions: First price, second price (vickery auctions), sequential auctions

11. Lecture 11: Agent learning I: Single agent learning (basics)
    Star Trek: The next generation episode: The offspring
    Homework: Read Asimov's "Little lost robot"for next lecture

Part II: Multiagent Interactions

12. Lecture 12: Agent modeling I: Symbolic plan recognition, model tracing, prediction

13. Lecture 13: Agent modeling II: Recursive agent modeling, Plan randomization for adversarial domains

14. Lecture 14: Biologically inspired multiagent systems, ant algorithms, emergent coordination

15. Lecture 15: Teamwork I: What is teamwork, team logic, mutual beliefs, joint t goals
    "Minority report"clip

16. Lecture 16: Teamwork II:
    a. Practically implementing teamwork beyond joint persistent goals: representing team plans and roles in an agent architecture, addressing practical communication costs, team monitoring and recovery from failures.
    b. Introduction to decision theoretic approaches to teamwork, distributed POMDPs.

    Homework: Vernor Vinge "Fast Times at Fairmont High"from "Hard SF Renaissance" for next lecture

17. Lecture 17: Intelligent agents field trip for demonstration

18. Lecture 18: Team formation (symbolic matching, combinatorial auctions), task allocation (contract nets), coalition formation
    Star Trek episode: Who watches the watchers
    Homework: Read Bruce Sterling "The Swarm" for the next lecture

19. Lecture 19: Distributed constraint reasoning, introduction to distributed constraint optimization (DCOP)

20. Lecture 20: Agent learning II: Multiagent learning
    Focus on the multiagent aspect of learning, e.g. in game contexts.

Part III: Agents and their impact on society

21. Lecture 21: Show initial clip of commander Data goes on trial; provide readings to students in preparation for the trial of commander data. Students divided into two groups, with each group divided into subgroups of 4 each, with each subgroup of 4 given one topic: (i) self-awareness and consciousness; (ii) rights and responsibilities; (iii)…
    Star Trek: The next generation "The measure of man"

    We will show the full episode of trial of commander data (a major part of it). Then we will form teams. We will assign readings. Readings will be photocopied from books or papers. Example readings include:
    - Searle's "Minds, brains and programs"
    - Nagel's "What it's like to be a bat?"
    - Turing's "Computing Machinery and Intelligence" – common counterarguments are presented.
    - Roger Penrose "Can a computer have a mind"

22. Lecture 22: Rights of agents: Students run trials
    Commander data on trial ends.

23. Lecture 23: Invited Lecture on Trial of commander Data

24. Lecture 24: Adjustable autonomy, Mixed-Initiative Planning, Decision theoretic approaches, strategies in adjustable autonomy
    Homework: *View one of:*

    "2001: A Space Odessey" or other movies or TV episodes, where agents or obots are presented in a negative light, and present a design (or a non-technical argument) for safety of agent based systems presented in class in a week's time.

25. Lecture 25: Safety in agent-based systems: Student presentations. Students should present either a technical design that avoids harm by the robot or AI system they chose to modify; or a non-technical argument as to how this harm could be avoided via societal modifications or new rules and regulations, or explain why this harm will never arise.

26. Lecture 26: Agents in the real-world: Massive, DS-1, ARMOR, Mix of applications.

27. Lecture 27: Wrapup discussion, Review, difference between science fact and science fiction