Social Agents

Norman Foo[1] and Pavlos Peppas[2]
[1]National ICT Australia and School of Computer Science & Engineering
University of New South Wales, Sydney, NSW 2052, Australia
Norman.Foo@nicta.com.au / norman@cse.unsw.edu.au
[2]Department of Business Administration
University of Patras, Rio, 26500 Patras, Greece
ppeppas@otenet.gr

ABSTRACT

Social agents are cognitive entities that reason about and interact with one another. Formal and computational models of social agents will have properties familiar to AI researchers, such as beliefs, desires, intentions, utility maximization, inferential mechanisms, etc., but in addition to these they will also have attributes, emotions, attitudes, dispositions, etc. that have hitherto been the domain of social scientists and psychologists. Moreover, complete rationality and deep reasoning have been shown to be unrealistic assumptions in human interactions and decision making in many contexts, so useful models of social agent communities should take these limitations into account. The roles played by bounded rationality and emotive attitudes cannot be sidelined. This position paper reviews work in this area and proposes a methodology for computational models, specifically using logic programs, of social agents that is amenable to both simulation and game-theoretic investigation. In particular we address the problem of social agent populations that are large, numbered in the thousands.

Keywords: social agent, limited rationality, emotion, computational model, logic program.

1. INTRODUCTION

The computational modelling of agents has a long history if one includes economics, game theory and artificial intelligence (AI) in it. In economics, questions about how a rational person should make decisions in environments where different decisions result in different payoffs were formalised and solved as early as the mid nineteenth century. A hundred years later game theory began to flower in the milieu of mathematical economics and heightenned international cold war tension. The period between the late fifties and the end of the twentieth century were very productive for the understanding of rational agents, and the incorporation of the insights from economics and game theory into AI. A key contribution of AI was the formalization of the cognitive apparatus of agents, making explicit the way agents reasoned about one another, and subjecting the processes to possible failure. Further, the notion of cooperation vs competition, often abstracted away in traditional approaches, received more attention. Much of this was driven by the need to ensure honest behavior in environments such as internet auctions, bandwidth bidding, etc. In parallel with these largely theoretical developments, questions arose about the reliability of the rationality assumptions of agents. This led to empirical research in what is now known as behavioral economics, where experimental studies have revealed that people make irrational decisions in many contexts. But more to the point, irrationality is not the same as randomness; the irrationality is systematic, and presumably amenable to scientific modelling.

This position paper reviews work relevant to social themes, and proposes a methodology for constructing computational models of agents that can be used to simulate and reason about behavior and large-scale properties of communities of these agents in social settings. In particular, it can accommodate systematic irrationality and emotive attitudes, and large agent populations. One motivation for this proposal is an understanding of the emergence of a number of attributes of current societies like mass anxiety, prejudice, aggression, etc. Another is an account of how the cuture of a corporation affects its performance. As this is only a position paper, there are few actual results; instead, it has a number of pointers to what we think should be done in the near future to realize this modelling.

2. REVIEW OF AREA

2. 1. Economic Theory

In economic theory approach to agent modeling the reasoning of each agent about itself and others are implicit. The models typically examine a single agent which has complete knowledge of its options, the

payoffs for each option, and assumes that all agents will
make choices that maximize their payoffs. This area had its beginnings in the mid eighteenth century, and is well reported in standard texts [1] on game theory as well as economics. But its most modern manifestations [2] try to resolve interesting social difficulties like the "free rider" problem that had its origin in the old "tragedy of the commons" phenomena that are in fact still relevant today. Nevertheless, these approaches have the features mentioned in the preceding paragraph, and are subject to all their limitations.

2. 2. Game Theory

The key ideas of modern game theory were crystallized in the seminal work of von Neumann and Morgenstern in which rational economic behavior under complete knowledge was modeled as a game. This approach became more convincing to practicing economics after Nash showed that mutual reasoning by interacting agents led to the *equilibria* (see [1]) named after him. The framework for specifying such games is the *normal form*, which is essentially a set of vectors, each of which encodes the agent payoffs for a possible *instantaneous and independent* choice of strategies by them. If one looks carefully at the reasoning used to establish the equilibria the first thing one notices is the *nested inference* that agents are supposed to perform, e.g., "He knows that I know that she knows …". Moreover this nesting is potentially unbounded. Unfolding this nesting *operationally* yields an alternative interpretation of games, that called the *extensive form,* the natural representation of which are game trees in which the nodes at various levels encode the turn of an agent to make a *move.* This is in fact the more intuitive form for games like chess and monopoly. There are known connections between the two forms.

2. 3. Artificial Intelligence

AI began borrowing from economics and game theory their notions of agents when it looked at ways to formalize multi-agency so as to render its models susceptible to execution and analyses. The work here has many facets, ranging from fairly standard ways for agents to update their beliefs [3] which hinges on rationality, to issues about the limitations imposed by bounded resources [4] and algorithms for cooperative planning. Understandably, the focus of the research has been dictated by requirements in areas such as network security, robotics and internet transactions. Social concerns like trustworthiness, credibility, compassion, altruism, etc. were at the periphery. However it is becoming clearer that they cannot be neglected forever. Trustworthiness, for instance, is at the core of reputation sytems like eBay's. There is not a lot of solid work on social agent properties comparable in depth and rigor to the economic and game theoretic literature on agent behavior, but one [5] has a qualitative overview of what affects trusworthiness while another [6] attempts a game-theoretic approach to this property. The former has the merit of extensive coverage, but that is also a weakness as any attempt to base computational models on it is doomed because of uncontrolled complexity. The latter has good suggestions on what must be essential to any formal model but falls short for the opposite reason of being too simple. However, it adopts a way of thinking about social properties that is inherited from the work of Axlerod on the evolution of cooperation.

3. A METHODOLOGY FOR SOCIAL AGENTS

Based on the preceding work reviewed in section 2 above we propose a methodology for modelling interacting social agents, initially just a small number (maybe only two) but later thousands or greater. The goal is to build models inspired by notions and techniques that are well-known but extending them in directions that can be used to test hypotheses about the modeling process itself. Further, we hope that the model classes that survive the process are sufficiently faithful to real world systems that their dynamics are helpful in prediction, explanation and validation.

3.1. Key Features for Models

There are several features that we believe are relevant for typical social agents. They are listed below for a and the details elaborated afterward.

a. Cognition
b. Reasoning
c. Emotion
d. Memory
e. Defaults
f. Adaptation

This is a selected list from work like [5], and an attempt to keep the complexity down. For a specific profile of social dynamics only a subset of these features may be needed. We now proceed to the elaboration.

Cognition. Agents not only think about their own plans of action but about others. Rather than keep this kind of thinking implicit in a model, it should be made explicit in some kind of epsitemic logic. The basic ideas can be imported from AI. A good way to test the formal models may be to encode traditional game-theoretic normal forms in them and see if the known equilbria can be easily captured. An indication of how this might be done is [7].

Reasoning. The main point to make here is that the models must allow for several feartures that are less than ideal, e.g., limited resources (bounded rationality), uncertainty, errors. Work in AI on limited resources and uncertainty, and in game theory on probabilistic versions of equilbria using Bayesian completions should provide examples on how to do this.

Emotion. This is where the insights from behavioral economics are useful. In a sense they are about irrationality and asymmetry, albeit of a systematic kind. We may need new modalities in formal agent models to represent emotive attittudes such as anger, vengefulness, compassion, suspiciousness, etc. There are vague hints from AI on how some this can be done, where the properties captured there are degrees of scepticism of agents.

Memory. This is not independent of Emotion but deserves a separate mention because it can be trivialised. People often remember past honesty on the one hand and broken promises on the other. This memory affects their future behavior, so the memory of past events is important. While it is true that all k-memory systems can be encoded into a Markov system, the encoding is not natural and would render the epistemic logic obscure. This suggests that Binmore's idea [8] that equilibria are alternatively thought of as the convergent point of iterated games can be imported to model emotive attitudes as they evolve according to agent interaction.

Defaults. For a given environment social agents often "take for granted" certain social norms. These norms can be modelled as background default rules, well-known in AI, that vastly reduce computational load. Conversely, when defaults are removed agents have to divert resources to secure things that had previously been there for "free". The selection of which background defaults apply for a social environment is part of the modelling process.

Adaptation. Social agents are evolving entities, meaning that computational models of them must have the capacity to change rule sets. This is the core of the AI area of program updates, and ideas from there should provide pointers on how to do this.

3.2. Formal Models

Taking into account the requirements above, a good candidate for a social agent modelling language is *Extended Logic Programs (ELP)* [9], enhanced with epistemic literals as suggested by Gelfond and has been used for updates.. Without going into details, among the merits of this recommendation are: it allows both default negation and classical negation, it has an elegant semantics, interactions between agents for the specific case of negotiation has been modelled in it [10], as also is the case with Nash equilbria [7], two open-source computational systems are available for computing such programs, and last but not least the prospects for tracking the dynamics of agent interaction using this formalism are good. Moreover, it is not hard to place constraints on agent use of resources in this formalism, and there are already probabilistic versions.

What about the modelling of thousands of agents? This is necessary for the simulation of emergent properties of entire societies. Although each agent may still be an ELP as above, the modelling of very large numbers of interacting agents should borrow ideas from evolutionary game theory. This proposal is not particularly novel. Sociobiology has shown how some commonplace observations about animal behavior is consistent with evolutionary game-theoretic equilibria. Surely there are many human counterparts that can be similarly accounted for, provided the models of individuals are more sophisticated, in other words less driven by instinct and more by reflection or attitudes.

REFERENCES

[1] J. Watson, *Strategy: An Introduction to Game Theory*, W.W. Norton, 2002.

[2] E. Clarke, *Demand Revelation and the Provision of Public Goods,* iUniverse Publishers, 2000 (reprint).

[3] P. Gardenfors, *Knowledge in Flux,* MIT Press, 1988.

[4] J. Rosenschein and G. Zlotkin, *Rules of Encounter Designing Conventions for Automated Negotiation among Computers*, MIT Press, 1994.

[5] R. Falcone & C. Castelfranchi, "Social Trust: A Cognitive Approach", In Castelfranchi C. & Tan Y.H., Trust and Deception in Virtual Societies (55-90). Kluwer Academic Publishers.

[6] P. Dasgupta, "Trust as a Commodity", in *Trust,* ed D. Gambetta, Basil Blackwell, Oxford, 1988.

[7] N. Foo, T. Meyer and G. Brewka, "LPOD Answer Sets and Nash Equilibria", Proceedings of the 9th Asian Computer Science Conference, Chiang Mai, 2004, Springer LNCS 3321, ed. M. Maher.

[8] K. Bimore, *Essays on the foundations of game theory*, B. Blackwell, 1990 **.**

[9] M. Gelfond and V. Lifschitz, "Classical negation in logic programs and disjunctive databases", New Generation Computing, pp. 365-385, 1991.

[10] Y. Zhang and N. Foo, "A Unified Framework for

Representing Logic Program Updates", Proceedings of the 20th National Conference on Artificial Intelligence, AAAI'05, Pittsburgh, July 2005.