Modelling Socially Intelligent Agents in Organisations

Bruce Edmonds

Centre for Policy Modelling
Manchester Metropolitan University,
Aytoun Building, Aytoun Street
Manchester, M1 3GH, UK
http://www.cpm.mmu.ac.uk/~bruce

Abstract

Some work on modelling boundedly rational agents in organisations is described. It is then argued that social intelligence is not merely intelligence plus interaction but should allow for individual relationships to develop between agents. This means that, at least, agents must be able to distinguish, identify, model and address other agents, either individually or in groups; in other words that purely heterogeneous interaction is insufficient. Two example models are described, the second in detail where agents act and communicate socially, where this is determined by the evolution of their mental models. Finally some problems that arise in the interpretation of such simulations is discussed.

Modelling Agents

At the Centre for Policy Modelling (CPM) we are interested in modelling real agents, which can be people or other institutional units (such as firms or departments). We do this by modelling these as intelligent software agents. This perspective means we have slightly different concerns than those concerned with designing agents or robots to meet particular goals. In particular we seek veracity over efficiency.

Thus we do not model using reactive agents, since a principal concern of ours is how the nature and development of the agents' internal models as they interact with other agents and its environment (see also the reasons in the section entitled Social Intelligence and Complexity). We take the strategy of explicitly representing the agents internal models in a specified language - usually of a quasi-logical or functional variety. This explicit representation makes it possible to limit, examine and analyse the agents models.

The agents we model have distinct limitations of resources - they are boundedly rational in several respects. They have limited memory, a limit on searches for improved models and a limit of their ability to make inferences from their models. Following what is known about real agents we ensure that their search for new models is incremental, rather than global in nature. The current memory, especially the agent's stock of current models, encodes a sharp path-dependency. The nature and framework of this is described in (Edmonds and Moss 1996; Moss and Edmonds forthcoming).

One particular technique we use is an adaption of the genetic programming (GP) paradigm (Koza 1992). Here the internal models belonging to the agents are held as a set of tree-based expressions. The selection among these is based upon their past predictive success or some endorsement-based mechanism. However, unlike standard GP, we do not always use the cross-over operator to introduce variation as this is irrelevantly global in operation, but prefer other operators such as generalisation and specialisation (Edmonds and Moss 1997).

Modelling Organisations

One of our chief concerns in modelling agents is to capture some aspects of organisational behaviour. We do this by modelling them as populations of interacting agents in a given structure. We do not necessarily do this down to the level of individual persons but sometimes stay at the level of departments or even whole firms (if they are themselves interacting). Here a theme we are investigating is the contrast between the official (usually hierarchical) structure of the firm and the unofficial structures that emerge as the individuals are frustrated by the formal structure.

In order to study such organisational models, we have developed a modelling language called, SDML – a Strictly Declarative Modelling Language (Edmonds, Moss and Wallis 1996; Wallis, Edmonds and Moss, 1995). This allows the flexible and theory-free modelling of composite agents in a declarative framework with object-orientated features. In particular this is particularly suited for modelling organisations built up in many levels in a composite manner - it allows for more better structured organisational models involving more complex cognitive agents, than some other systems (Moss et al. 1996).

Modelling Social Interaction

Communication is a special case of action and perception, and in many organisational models the communication is very rudimentary. However communication is so important in social situations and potentially so computationally onerous that effectively it becomes a separate consideration.

One concern, in this regard is the passivity of the communication in most models of interaction. One gets the impression that in many models agents tend to make requests for information and sometimes issue orders but will not, for instance, volunteer unrequested information. Thus most models of communication correspond to a mutual (or merely unidirectional) *pull* of information rather than using a mix of *push* and *pull* modes.

Social Intelligence and Complexity

Social intelligence implies more than mere interaction with other agents plus intelligence. For example, agents might apply their intelligence to trading with other agents without any intelligence being applied to the *process* of relating to the others. Such an intelligence might be without the means of recognising and referring to other agents as individuals (or groups). In such a case its internal models of its social environment would be entirely generic and thus it could not form any social relationships with other agents (other than with the population as a whole). Such a lack of social intelligence has advantages, such as the ability to analyse and predict their behaviour in computational communities. However if we are to model many key behaviours in organisations, we need a greater social sophistication.

Thus in denoting the presence of a social intelligence, I would want to be able to identify at least some of the following:

- a relative sophistication of communicative mechanisms;
- the ability to represent aspects of other agents (individually or grouped), in order to anticipate their actions (though this need not involve the *explicit* representation of the other's beliefs, goals etc.);
- the ability to distinguish between and refer to different agents, such that different aspects may be captured for each one (or each group), e.g. their individual reliability as an information source;
- the presence of purely communicative (social) sub-goals (or even top goals).

In addition to these I think there is another aspect to social intelligence - that of dealing with the great complexity that social systems typically produce (which seems to grow exponentially with size, Carneiro 1987). In fact it seems to be a hallmark of social systems that such complexity arises due to the variety of individual specialisations and hence relationships that can develop. A society consisting only of heterogeneous agents will not have the same characteristics.

Luhman has argued that one of our social institutions' primary functions is to *filter out* the complexity of the external social world (as summarised in Bednarz 1984). This perspective highlights some other important aspects of social intelligence, including:

- the intelligent but restrictive selection of information sources;
- the development of rules to structure social interaction - either formally or informally (e.g. emergent social norms);
- the development of binding long-term relationships (contracts, friendships, etc.).

A socially intelligent agent may thus seek to use institutions which deal with the complexity of social reality by preforming considerable selection, and modelling for it and also by regulating and hence simplifying the social structure within. Such an institution may itself embed itself within a further institution.

Example 1 - A Model of Emerging Markets

A three-sector model of emerging market economies has been developed where the component firms leant how to behave in a newly emerging market economy (and in particular how to respond to the debt crisis). Firms communicate and trade with other firms and build rule-based models of their environment and other firms. Known qualitative behaviour characteristic of such economies (e.g. Belarus) was only reproduced when the model was adapted so that firms *only* copied the observable behaviour of other successful firms they interacted with. Here the ability to distinguish, identify, and select other firms was critical on the overall behaviour. A report on an early version of this model can be found in (Moss and Kuznetsova 1996).

Of course, this model only starts to address the concerns I listed above about social intelligence, but it does show how the inclusion of some elements of social intelligence can critically effect the emergent properties of the whole system.

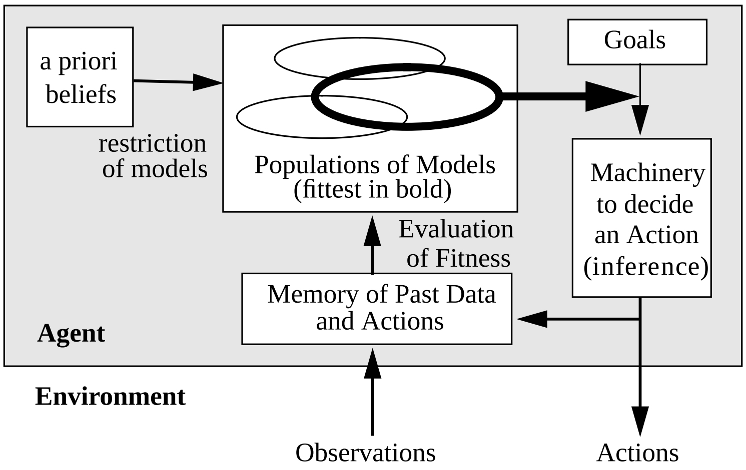

Figure 1: Basic structure of an agent

Example 2 - the El Farol Bar Problem

Description

I have extended Brian Arthur's El Farol Bar model (Arthur 1994) to include model-based learning and communication. In this example a population of agents has to decide whether to go to El Farol's each thursday night. It is generally desirable to go but not if it is too crowded. I have extended this by adding a social structure. A randomised "acquaintance" structure is imposed upon the agents, limiting who they talk to. Agents have a chance to communicate with acquaintances before making their decision. Agents have a population of models of their environment which is composed of a pair of expressions: one to determine the action (whether to go or not) and a second to determine their communication with other agents. Either of action or communication can be dependent upon communications received, which includes the identity of the agent the communications were received from. These internal models are developed using feedback from their experiences resulting from their decisions. Although the beliefs and goals of other named agents is not explicitly represented, they emerge implicitly in the agents' models.

The agent modelling approach broadly follows (Edmonds and Moss 1997). Thus each agent has a population of mental models, which broadly correspond to alternative models of its world. This population develops in a slow evolutionary manner based either on the past accuracy of the models predictions or some measure of what its past success at gaining utility might be. The agent structure is shown in figure 1.

Each notional week, the new population of models is produced as in a genetic programming manner (Koza 1992) using some tree crossover but with a high degree of propagation and also some new random genes introduced each time. Then the best model is selected and used to determine first its communicative action and subsequently whether to go to El Farol's or not. Thus the evolution of mental models is a rough representation of learning. The cross-over operation is very realistic but does as a first approximation, for a critique of cross-over for these purposes, see (Edmonds and Moss 1997).

Each model is composed of two parts: one determines what it says and the other what it decides. These parts are expressions from a two-typed language set at the start. A simple but real example model is shown in figure 2 below. Translated this means: that it will tell its 'friends' that it will go to El Farol's if the trend predicted over observed number going over two weeks is greater than 5/3 (the total population was 5 in this example); but it will actually go if it said it would go *or* if barGoer-3 said it will go.

talk: [greaterThan [trendOverLast [2]] [divide [5] [3]]]

action: [OR [saidBy ['barGoer-3']] [ISaid]]

Figure 2: A simple example model

```
talk-1: averageOverLast(numWentLast) > previous(trendOverLast(numWentLast))
action-1: wentLastTime

talk-2: trendOverLast(numWentLast) - 2 * numWentLag(2) > numWentLag(numWentLast)
action-2: NOT Isaid

talk-3: randomNumberUpTo(8) < 8/3
action-3: True

talk-4: averageOverLast(4)/averageOverLast(5) < numWentLag(15)
action-4: (Isaid AND randomDecision) OR (saidBy 2)

talk-5: trandOverLast(20) < numWentLag(2) - averageOverLast(numWentLast)
action-5: randomDecision OR (saidBy 4)
```

Figure 3: Simplified talk and action genes for the five agents at date 100

The agent gains utility by going to the El Farol Bar when it is not too crowded. Thus each agent is competitively developing its models of what the other agents are going to do.

A Case study from the results

Graphs of the results do not show anything surprising. The average fitness of the agents models fluctuate wildly at the beginning with a big difference between agents but as the simulation progresses they all settle down to around the same value. The deviance between different models of the same agent also reduces. When you look at the pattern of who goes and who does not, some of the agents settle for a fixed strategy but some are more dynamic and constantly swap between different strategies and elaborate old ones.

What is perhaps more revealing is the detail of what is going on, so I will exhibit here a case study of the agents at the end of a simulation.

Here I have chosen a 5-agent simulation at date 100. Inn this simulation the agents judge their internal models the utility they would have resulted in over the past 5 time periods. This utility function that agents get is 0.4 if they go when it is two crowded, 0.5 if they stay at home and 0.6 if they go when it is not too crowded (where too crowded means greater than 60% of the total population). This is supplemented with an extra 0.1 of utility for every one of their friends that go if they do.

The friendship structure is chosen at random at the beginning, and in this case is as show in figure 4 below.

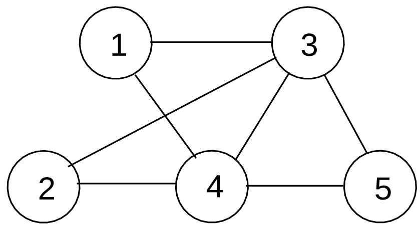

Figure 4: Friendship structure

The best (and hence active) genes of each agent are summarised above in figure 3. I have simplified each so as to indicate is *logical* effect only. The actual genes contain much logically redundant material which may put in an appearance in later populations due to the activity of cross-over in producing later variations.

The effect of the genes is tricky to analyse even in its simplified form. For example agent-1 will tell its friends it will go to El Farol's if the average attendance over a previous number of time periods equal to the number who went last time is greater than the predicted number indicated by the trend estimated over the same number of time periods but evaluated as from the previous week! However its rule for whether it goes is simpler - it goes if it went last week.

You can see that for only one agent does what it says indicated what it does in a positive way (agent 4) and one which will do the exactly the opposite of what it says (agent 2). It may seem that agents 1 and 3 are both static but this is not so because figure 3 only shows the fittest genes for each agent at the moment in terms of the utility they would have gained in previous weeks. During the next week another gene may be selected as the best.

The interactions are summarised in figure 5, which shows the five agents as numbered circles. It has simple arrows to indicate a positive influence (i.e. if agent 2 says she is going this makes it more likely that agent 4 would go) and crossed arrows for negative influences (e.g. if agent 2 says she will go this makes it less likely she will go). The circles with an "R" represent a random input.

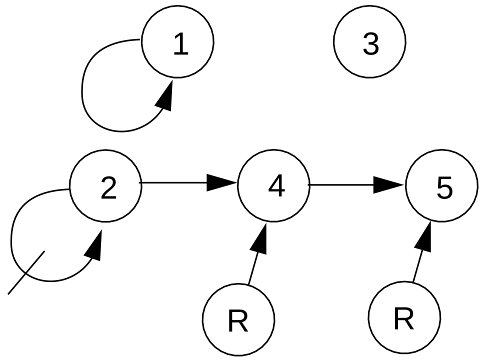

Figure 5: Talk to action causation

It is not obvious from the above, but agent-2 has developed its action gene so as to gradually increase the number of 'NOT's. By date 100 it had accumulated 9 such 'NOT's (so that it actually read NOT [NOT [... NOT [Isaid]...]]). In this way it appears that it has been able to 'fool' agent-4 by sometimes lying and sometimes not.

Issues and Interpretations

There are some very interesting problems that arise when we try to interpret what occurred. Even given that we can look inside the agents' heads one comes across some of the same problems that philosophers, psychologists and social scientists encounter in trying to account for human communication. The web of cause and effect can be very complex and so impede a straightforward analysis just as if the human case.

One issue in particular is the question of the "meaning" of the agent's utterances to each other. Their utterances do *have* a meaning to each other otherwise they would quickly select out action genes that included "saidBy" clauses. However, these meanings are not obvious. They are not completely determined by their own model structures, but can involve a number of language games whose ultimate grounding is to the practice of such communication in relation to actual decisions. Thus, in this particular example it seems that the pragmatics of the situation are the most important for determining meaning, followed by a semantics grounded in the effects of their actions, leaving the syntax to merely distinguish between the two possible messages. thus this case seems to illustrate Peter Gärdenfors observation about human language (Gärdenfors 1997):

> *"Action is primary, pragmatics consists of the rules for linguistic actions, semantics is conventionalised pragmatics and syntax adds markers to help disambiguation (when context does not suffice)."*

Conclusion

Although many of the suggestions about what a socially intelligent agent might involve where not included in the implemented examples, they show that even if we equip our agents with some of the tools necessary for truly social behaviour many aspects of real social behaviour can emerge.

When such social behaviour does occur we may well find ourselves with many of the same difficulties that other social scientists have, namely the tracing of very complex chains of causation if one works in detail and the problem of the meaning and use of our descriptive terms if we attempt a macroscopic approach.

Acknowledgements

We acknowledge the support of ParcPlace-Digitalk for providing the VisualWorks environment in which SDML has been developed.

References and Related Work

(*All 'CPM' reports can be accessed from URL:* http://www.cpm.mmu.ac.uk/cpmreps.html)

Arthur, B. 1994. Inductive Reasoning and Bounded Rationality. *American Economic Association Papers*, 84: 406-411.

Bednarz, J. 1984. Complexity and Intersubjectivity. *Human Studies*, 7: 55-70.

Carneiro, R. L. 1987. The Evolution of Complexity in Human Societies and its Mathematical Expression. *International Journal of Comparative Sociology*, 28: 111-128.

Edmonds, B. and Moss, S. J. 1996. The Credible Modelling of Economic Agents with Bounded Rationality. In Ein-Dor, P. (ed) *Artificial Intelligence in Economics and Management*. Boston: Kluwer Academic, 205-21.

Edmonds, B., Moss, S. J. and Wallis, S. 1996. Logic, Reasoning and a Programming Language for Simulating Economic and Business Processes with Artificial Intelligent Agents. In Ein-Dor, P. (ed.): *Artificial Intelligence in Economics and Management*. Boston: Kluwer Academic, 221-230. Also report CPM-96-09.

Edmonds, B. and Moss, S. J. 1997. Modelling Bounded Rationality using Evolutionary Techniques. *AISB'97 workshop on Evolutionary Computation*, Manchester. To be published in the Lecture Notes in Computer Science, Springer, Berlin. Also report CPM-96-10.

Gärdenfors, P. 1997. The pragmatic role of modality in natural language. 2*0th Wittgenstien Symposium*, Kirchberg am Weshel, Lower Austria, 1997.

Koza, J. R. 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press.

Moss, S.J. and Edmonds, B. 1997. Modelling Economic Learning as Modelling. *Systems and Cybernetics*. Forthcoming. Also report CPM-94-03.

Moss, S. J., Gaylard, H., Wallis, S. and Edmonds, B. 1996. SDML: A Multi-Agent Language for Organizational Modelling. Submitted to *Computational and Mathematical Organization Theory.* Also report CPM-97-16.

Moss, S. J. and Kuznetsova, O. 1996. Modelling the Process of Market Emergence. In Owsinski, J. W. and Nahorski, Z. (eds.). *Modelling and Analysing Economies in Transition*, MODEST, Warsaw, 125-138. Also report CPM-95-12.

Wallis, S., Edmonds, B. and Moss, S. J. 1995. The Implementation and Logic of a Strictly Declarative Modelling Language. In Macintosh, A. and Cooper, C. (eds.). *Applications and Innovations in Expert Systems III*. Oxford: SGES Publications, 351-360.