Lecture Notes in Artificial Intelligence 1979

Subseries of Lecture Notes in Computer Science
Edited by J. G. Carbonell and J. Siekmann

Lecture Notes in Computer Science
Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Tokyo

Scott Moss Paul Davidsson (Eds.)

Multi-Agent-Based Simulation

Second International Workshop, MABS 2000
Boston, MA, USA, July
Revised and Additional Papers

Springer

Series Editors

Jaime G. Carbonell,Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Scott Moss
Manchester Metropolitan University
Centre for Policy Modelling
Aytoun Building, Manchester M1 3GH, UK
E-mail: s.moss@mmu.ac.uk

Paul Davidsson
University of Karlskrona/Ronneby
Department of Software Engineering and
Computer Science
372 25 Ronneby, Sweden
E-mail: Paul.Davidsson@ipd.hk-r.se

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Multi-agent- based simulation : second international workshop ; revised and
Additional papers / MABS 2000, Boston, MA, USA, July. Scott Moss ;
Paul Davidsson (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ;
London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2001
(Lecture notes in computer science ; Vol. 1979 : Lecture notes in
artificial intelligence)
ISBN 3-540-41522-X

CR Subject Classification (1998): I.2.11, I.2, C.2.4, I.6, J.4, H.5

ISBN 3-540-41522-X Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

Printed in Germany

Typesetting: Camera-ready by author, data conversion by DA-TeX Gerd Blumenstein
Printed on acid-free paper SPIN 10781276 06/3142 5 4 3 2 1 0

Preface

This volume is based on papers accepted for the Second International Workshop on Multi-agent-based Simulation (MABS-2000) federated with the Fourth International Conference on Multi Agent Systems (ICMAS-2000) held in Boston in July 2000.

The purpose of MABS-2000 was to investigate and develop the synergy between software engineering for multi-agent systems and agent-based social simulation. The papers included in the MABS-2000 workshop were selected either because they explore how agent interaction can be used to build multi-agent systems or they offer examples of problem-oriented (rather than technique-oriented) systems. No paper was selected if it specified a model or an issue to make it fit a previously chosen technique.

All of the papers in the volume have been reviewed and in many cases revised since the workshop. Two papers (by Edmonds and by Hales) as well as the editorial introduction have been added to those accepted for the workshop.

As editors and workshop organisers, we are very grateful to the participants who engaged enthusiastically in the discussions about both individual papers and the issues facing the MABS community. Issues raised and positions taken in those discussions are reported in the editorial introduction. We are also grateful to the authors for their punctuality and the grace with which they received and responded to editorial comments and requests. Klaus Fischer, the ICMAS-2000 workshops chair, was exceptionally patient and diplomatic in reconciling our demands with the resources available. We are particularly grateful to Klaus and to Ed Durfee, the general chair of ICMAS-2000, for arranging the infrastructure for an extra day of MABS-2000 when it became clear that that was warranted by the number and quality of papers submitted.

A special word of thanks is due Jaime Sichman, an editor of the proceedings of MABS'98. Jaime gave us very useful, timely and welcome advice on the organisation and planning of the workshop and in organising the publication of this volume. Finally, we thank Alfred Hoffmann and his team at Springer-Verlag for enabling us to combine the further reviewing, extension and revision of papers with rapid and timely publication of the volume.

November 2000

Scott Moss
Paul Davidsson

Programme Committee

Rob Axtell (The Brookings Institution, Washington)
Eric Baum (NEC Research Institute, Princeton)
Magnus Boman (Stockholm University)
Rosaria Conte (IP-CNR, Rome)
Kerstin Dautenhahn (University of Reading)
Keith Decker (University of Delaware)
Jim Doran (University of Essex)
Bruce Edmonds (Manchester Metropolitan University)
Nigel Gilbert (University of Surrey)
Jaime Sichmann (University of São Paolo)
Klaus Troitzsch (University of Koblenz)

General Chair

Scott Moss (Manchester Metropolitan University)

Co-chair

Paul Davidsson (University of Karlskrona/Ronneby)

Table of Contents

Editorial Introduction: Messy Systems – The Target for Multi Agent Based Simulation

Scott Moss

Centre for Policy Modelling
Manchester Metropolitan University
Manchester M1 3GH UK
s.moss@mmu.ac.uk

Abstract. Messy systems have no clear boundaries; they are composed of so many natural and/or artificial entities with patterns of interaction so dense that they cannot be understood by inspection and system behaviour cannot be predicted by statistical or qualitative analysis. Obvious examples are real social systems and the Internet. Analysing and designing software to exploit such systems requires a different approach to software engineering and mechanism design. The issue addressed in the MABS-2000 workshop and in this volume is the development of a methodology and technology to identify which techniques hold promise and which cannot possibly lead to useful applications for messy software or social systems.

1 Messy Systems and Tidy Systems

Much of AI and the great preponderance of papers at conferences such as ICMAS relate to software systems with clear boundaries and well understood relationships among known entities. The perceived development path is by means of step-by-step development of well understood formalisms and algorithms. These are tidy techniques for tidy systems.

Real systems are frequently messy. The boundaries of action and opportunity are unclear and probably unknowable. Relationships among entities are either too complex to be understood or shifting faster than we can identify them. In short, we frequently do not know enough about the systems to make them amenable to tidy techniques for toy systems.

The differences between messy and tidy systems have been recognised for some considerable time by the agents research community.

Wooldridge and Jennings [20] argued that "[i]f a system contains many agents…, then the dynamics can become too complex to manage effectively. There are several techniques that one can use to try to manage a system in which there are many agents. First, one can place it under central control... Another way... is to severely restrict the way in which agents can interact... one can ensure that there are few channels of communication... [or] by restricting the way in which agents interact. Thus very

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 1-14, 2000.

simple cooperation protocols are preferable…." This quotation is taken from Edmonds [8] who commented that "these are sensible warnings for the software engineer, but they are not necessarily relevant for social simulation, since the unforeseen behaviour that the engineer is trying to prevent is what the social simulator is interested in. For the social simulator, the issue of how society can impact upon individual behaviour is at least as important as how individuals impact on society."

More recently, we have observed the growing articulation of an alternative view stemming from the increasing importance of such messy distributed software systems as the Internet and large, federated databases. While the importance of agents in software engineering research is hardly in doubt, there have been some rumblings of disquiet about the scalability and breadth of applicability of the agents paradigm. A classic paper by Nwana and Ndumu [15] points out a range of problems that the agents community has largely ceased to address – much less to resolve and concludes that

> A new field is *only* defined by its *problems*, not its *methods/techniques*. We argue strongly that MAS has to some degree been falling into the trap that … befell AI – that of deluding itself that its methods and techniques (e.g. cooperation, rationality theories, agent languages, conceptual and theoretical foundations, multi-agent planning, negotiation) are the real important issues. No! They are not! It is the problems that they are meant to solve….[1]

Another way of putting the same point is to observe that the MAS research community is turning from difficult, messy problems in order to address *possibly* inconsequential issues that are certainly technically challenging. These issues include the development of formal logics, game theoretic strategies or various other optimisation techniques. A nice example of what happens here is from the ICMAS-2000 paper by Brainov and Sandholm [2]:

> Reasoning about others and interactive knowledge have been the subject of continuous interest in multiagent systems …, artificial intelligence … and game theory. In multiagent interaction, where an agent's action interferes with other agents' actions, hierarchies of beliefs arise in an essential way. Usually an agent's optimal decision depends on ….

Brainov and Sandholm then devise a means of representing infinite belief hierarchies (what I believe you believe about what I believe you believe … about me) as finite graphs. They then offer an example of how this approach supports the analysis of auctions. The standard assumptions, as described by the authors, are:

- Bidders are risk neutral
- Payment is a function of bids alone
- The auction is regarded in isolation from other auctions
- Bidders private valuations are independently and identically distributed random variables

[1] MAS is by no means the only research field where technique sometimes seems to prevail over problems and applications. Moss [10] documented that specific core techniques in economic theory are so important that formal demonstrations of their invalidity are repeatedly and systematically ignored. *Cf.* Moss and Pahl-Wostl [12] or Moss, Pahl-Wostl and Downing 13].

- Every bidder knows his own valuation
- Bidders share common knowledge about the distribution from which valuations are drawn

Brainov and Sandholm drop the last of these assumptions. The power of replacing the common knowledge assumption by the finite belief graph representing an infinite belief hierarchy is demonstrated by an auction in which there are two risk-neutral bidders for a single indivisible object, each buyer has one of the only two *possible* valuations of the object, just before the beginning of the auction the distribution of *objective* valuations has changed, each bidder has observed the change but neither knows the other has observed it. This, the authors assure us, is a realistic assumption for electronic auctions.

No evidence is actually provided for the realism of any of the original six assumptions or for the realism of the example auction specified in the paper. All that can be stated for certain is that Brainov and Sandholm have analysed the consequences of substituting the assumptions required to obtain their finite belief graph for the common knowledge assumption. Whether the substitutions are more realistic than the original assumption is, at best, an open question.

The fundamental point here is that the authors' formulation of the issue they addressed and their assessment of it was entirely within the closed world of game theory, decision theory and formal logics. Claims of realism were made without any evidence. The example given served to verify but not to validate the technique.

Though perhaps an extreme case, the Brainov-Sandholm paper serves nicely to exemplify the practice of developing technique with hope but no analysis of the ultimate applicability of the technique in messy auction environments. At a minimum, real bidders are influenced by social context, they know that there are other auctions and valuations are not realisations of i.i.d. random variables. Moreover, bidders are sometimes uncertain about their own valuations and are influenced by, for example, the pace of the bidding and the rate of change of the bids. If any or all of these considerations are not relevant, evidence of their irrelevance has not been offered.

The issue to be addressed – and which was addressed in the MABS-2000 papers and workshop – is the development of a methodology and technology to identify which techniques hold promise and which cannot possibly lead to useful applications for messy software or social systems.

2 Agents for Tidy Systems

The point of working with formal specifications and tidy models tested on toy systems is to build step by step towards useful applications where at least some of those applications could be used in the messy worlds of electronic commerce, real electronic auctions, data mining in large, complex databases, robotics for open environments, and so on. Such research is intended to provide a foundation for software engineering for messy systems. In the agent based social simulation (ABSS) community, this foundational approach is complemented by representational modelling – that is, by the implementation and testing of models intended to capture important aspects of actual social systems which are, of course, inherently messy.

There is a similar complementarity between foundational approaches in the wider MAS community and software engineering for tidy systems. Intelligent manufacturing and the development of optimisation techniques is an example of an area in which this complementarity is productive. A good example from ICMAS-2000 is the paper by Bussmann and Schild [3]. They had a clear manufacturing control objective for which they devised, and formally proved the properties of an auction protocol. A robust, formally sound, agent and mechanism design was devised to meet a specific engineering need and resulted in an agent based software application.

The Brainov-Sandholm and the Bussmann-Schild papers are opposite and extreme examples of the relationship between useful application and formal analysis. The history of science and technology indicates the Bussmann-Schild approach to be the more productive. A number of contributors to the historical literature on technology and organisation have concluded independently that the development of technology stems from the focusing effects of constraints. Some constraints are typically seen as the most binding in the sense that the payoff to the shifting of those constraints would be greater than from the shifting of any other constraints. Numerous instances of this phenomenon have been documented by Rosenberg [18] in the development of manufacturing technology and by Penrose [17] and Chandler [5] in the development of systems (including organisation) technology. The difference between the extreme formalism of the Brainov-Sandholm paper and the applied formalism of the Bussmann-Schild paper is in the source of the constraints that each is seeking to overcome and the criterion of success. Brainov and Sandholm are addressing a constraint that is internal to the framework of their analysis while Bussman and Schild are addressing a common constraint on manufacturing throughput. Success for Brainov and Sandholm is to substitute a formal structure for a widely used assumption. They assert without evidence that this is more plausible but the only criterion of success is the authors' estimation of plausibility. Success for Bussman and Schild was the demonstration of a prototype manufacturing line that exhibited the properties indicated by the formal analysis.

Although doubts are being raised here about the value of the sort of extreme formalism reported by Brainov and Sandholm, the development of formalisms without immediately direct application in principle can provide a toolkit from which applications are developed. Formalisms applied to a sequence of test problems, each test incorporating some further aspect of real applications, is common and sound engineering practice. An example, also from ICMAS-2000, is the paper by Sullivan, Grosz and Kraus [19] reporting an agent design allowing for flexibility in the meeting of commitments to common objectives. The problem addressed in that paper is the ability of software agents to respond appropriately to environmental factors without creating the retaliation problems encountered in game theoretic formulations. One example offered by the authors is the agent responsible for backing up files defaulting on that responsibility in order to assist on a crash recovery. Although much of the argument is couched in terms of agent utilities and utility maximisation, in fact the authors drive a coach and horses through formal utility theory by emphasising the utility individuals get from the utilities of other individuals, and resolving scheduling issues by means of a scoring system bearing more than a passing resemblance to Paul Cohen's [6] endorsements schemes.

While Sullivan *et al.* are by no means reporting applications for messy systems, the agent and mechanism designs are certainly driven by non-formal problems capturing aspects of applications requirements. Moreover, although Sullivan *et al.* set the context of their paper in terms of a clear progression from previous work, their simulation techniques were developed to meet a problem they were having in getting software agents to reconcile their "intensions" so they could function as collaborative partners. As Grosz put in a personal communication[2], "We wanted to deal with the [intention reconciliation] problem as it was likely to occur. We are gradually building from the simplest scenarios to more complex ones. It's very hard to do, harder even than I thought. However the techniques are in the service of the problem."

Science and technology develop iteratively: problems focus attention on the need for new concepts and technology while new concepts and technology create opportunities to address new problems, The particular problems range from the directly applied to the analytically introspective. History does not suggest that highly formal disciplines, such as the economic theory used by Brainov and Sandholm, provide useful insights or applied technologies. In the case of economics of the sort used by Brainov and Sandholm, it is arguable that the last unambiguously useful theoretical development was David Ricardo's law of comparative advantage published more than 180 years ago. While we cannot prove that the economics based approaches never will lead useful applications, to believe that they will do so requires a much more substantial act of faith than the more rough and ready (though still formally sound) approaches reported in the papers by Bussman and Schild or Sullivan, Grosz and Kraus.

3 MABS: Agents for Messy Systems

The lessons of both history and the experience of agent research for tidy systems suggest that software engineering for messy systems will usefully focus on the development of techniques for real systems and problems and representations thereof. The messiness of the real target systems implies considerable agent interaction. In social simulation modelling, agent interaction is evidently part and parcel of any representation of social processes and institutions. The need for agents to interact even in quite tidy systems is generally acknowledged by the agent research community if only implicitly through the number of papers on collaborative and cooperative agents. Frequently, means are found to mediate the interaction via auctions and, sometimes, brokers. However, Moss [14] has developed social simulation models to demonstrate that intermediated action is not always efficient or appropriate. The conditions in which they are efficient and appropriate elements of mechanism design are drawn from historical analysis and applied with some success to the analysis of brokerage by agents in information filtering mechanisms.

The whole issue of the conditions in which agent behaviour will lead to observed or intended system results is a thread running through the fabric of agent based social simulation. In general, these conditions are tied up with issues of validation. For social simulation modelling, validation turns on observed correspondences between agent behaviour and the behaviour of the social actors those agents represent and also between the macro behaviour of the model system and the observed behaviour of social institutions and the evidence of social processes. An important issue that

[2] Reproduced here with permission.

exercised the open discussion sessions of MABS-2000 was the extent to which validation should turn on prediction or description. Although there was considerable – majority – sentiment in favour of validation by prediction, a reasonable reading of the relevant papers at the workshop and in this volume indicates that validation in practice compares qualitative and quantitative model outputs with descriptions from the scientific literature and prior observations of statistical time series.

The agent based social simulation community has long been interested in interaction among software agents both as representations of human and social actors and to formalise elements of agent design for multi agent software systems. Major contributors to this tradition are Castelfranchi, Conte and their collaborators from IP-CNR in Rome. Classic examples are Conte and Castelfranchi [6] and Castelfranchi and Falcone [4] which use formal logic to investigate the meaning of such concepts as social structure or trust and how these emerge from and influence agent interaction.

The papers in MABS-2000 either explore how agent *interaction* can be used to build multi agent systems or they offer examples of problem-oriented (rather than technique-oriented) systems. In keeping with the mainstream of multi agent based simulation, no paper simply specifies a model or an issue to make it fit a previously chosen technique.

Three papers that draw explicitly and directly on sociological theory are included in the present collection. Two (Sawyer and Schillo *et al.*) are position papers addressing key issues in the development of multi agent systems and the third (Pedone and Conte) formalises the Simmel hypothesis on social differentiation within a grid-based simulation model. Together with Axtell's investigation of the effects of spatial representations of relationships among agents and the sequencing of their actions, these four papers address interaction issues from a more abstract perspective than any of the other papers. It may be useful to offer a framework within which to relate these papers.

Writers as diverse as the sociologist Mark Granovetter [12] and philosopher/modeller Bruce Edmonds [8] have independently developed the concept of social embeddedness. Individuals are socially embedded if their decisions and actions cannot be understood except in a social context. Wooldridge and Jennings argued in the passage quoted above that, in effect, socially embedded software agents are less reliable in an engineering sense than are socially unembedded agents. Granovetter pointed out that all agents in economic theory are not in any way socially embedded since their decisions and behaviour are determined entirely by their utility functions. He also pointed out that in some approaches to sociology, behaviour is so determined by social forces that there is little or no freedom of action for the individual. In his terminology, economic agents are "undersocialized" while individuals in Parsonian sociological theory are "oversocialized".

The general point here is that the degree of socialization or, perhaps more clearly, the form and extent of social embeddedness ought to be chosen on the basis of experience and experiment. The point seems to have some force *a priori* whether we are concerned with agents to represent human actors or software agents to act independently in large, complex social environments.

Three of the papers in the workshop collection address precisely this issue. Sawyer offers an extended example of observed and documented social embeddedness in order to discuss the phenomenon of emergence. In his example, emergent behaviour is observed in an improvisational theatre performance. His paper is intended to clear the ground for the development of multi agent systems with properties that emerge clearly as consequences of social interaction among agents. While Sawyer was not concerned explicitly with the determination of an appropriate form of social embeddedness, the importance of the form (as distinct from some abstract *degree*) of social embeddedness is central to his discussion. Schillo *et al.* do address the issue of the degree of embeddedness, again from a sociological and abstract perspective. Axtell, though not concerned with the more abstract sociological issues deals with the representation of social interaction. He demonstrates in particular that the results from models implementing different representations of social interaction yield qualitatively different individual and social behaviours. Taking these papers together, we have the implication that the choice of the form and representation of social embeddedness is highly non-trivial. These results give further force to the argument that multi agent systems – whether as software engineering or as social simulation – should be designed with the problem in mind, recognising the potential importance of the form of social embeddedness for the validity of the system.

The five papers grouped under the applications heading are all examples of problem-driven multi agent simulation models. Davidsson extends the work of Parunak *et al.*, reported in the first MABS workshop, on the differences in the results obtained from multi agent simulations and those obtained from system dynamics and other more mathematical simulation models. The discussion is problem-centred in that the example used to develop the argument and render it more concrete is an actual problem of designing intelligent building systems. McGeary and Decker are working on mechanism and agent design for the scalable control and coordination of systems where "agents must interact in complex, changing ways." They report work on a simulation system for investigating the mechanism and agent designs to support such interactions where necessary for the effective performance of software systems. While McGeary and Decker are working within relatively abstract, general systems, El hadouaj *et al.* take a similar problem-centred approach to a highly specific, empirical issue. They consider the psychological evidence on driver behaviour in road traffic, how well such behaviour supports observed traffic patterns and then assess and develop their agent representations to conform to both observed patterns in traffic systems and the evidence on individual behaviour. While they do not claim complete success, they do report improvements based on validation procedures involving domain experts. A third discussion of alternative simulation approaches to address a problem issue is offered by Breton *et al.* who argue explicitly for the advantages of multi agent simulations over previous approaches to the analysis of the dynamic properties of piles of granular material such as sandpiles, grain in storage elevators and the like. In this paper, uniquely for the present collection, the agents are representing inanimate objects (grains of sand) rather than animate (usually human) creatures. The paper by Kafeza and Karlapalem investigates and extends the work flow management system framework to speed up activities in multi agent environments, using simulations to assess the effectiveness of their innovation. While this paper is the least clearly focused on agents, it is also the clearest in raising implementation issues that affect agent and mechanism design.

Three papers address issues of actual social interaction in order to develop clear hypotheses for empirical validation. Hemelrijk's paper on sexual attraction and dominance, builds on her own and others' earlier work to develop new hypotheses about the effects of aggressiveness, social cohesion and sexual attraction – several of these hypotheses emerging unexpectedly from the simulations. Rouchier *et al.* report and compare three models developed in the French research institute CIRAD that generate hypotheses concerning the use of common renewable resources. The issues addressed in all of the models reported in both of these papers were selected because of their empirical importance and lack of obvious resolution in their literatures. The implementation of each model was also guided and informed by empirical observation of relevant populations of primates and butterflies in Hemelrijk's paper and human communities in western Africa the paper by Rouchier *et al.* The third paper in this vein, by Downing *et al.*, outline a strategy for modelling large scale, complex environments and reports a proof-of-concept demonstrator developed and validated with the participation of domain experts. The particular application is the development of integrated physical-biological-social simulation models to inform the policy process concerning the mitigation and effects of climate change. The reliance on agent based social simulation models in such integrated policy assessment applications is novel and uses an agent based software engineering approach dramatically to push back the trade-off between the scale of the system to be simulated and the detail of the representation.

A fourth paper addresses issues of social interaction on a more abstract plane. This is Hales' paper in which he demonstrates conditions in which cooperation will emerge from social interaction. Hales models strategy evolution in the context of a prisoners' dilemma game. It is not, however, a game theoretic model since the agents are not given pre-determined strategies and they are not maximising their payoff or utility. Instead, each agent is defined by a bit string (its "tag") plus a further bit to determine whether the agent cooperates or defects. Each agent plays, if it can, an agent with an identical tag and defects or cooperates according to the value of its "strategy" bit. After every agent has played, there is a genetic-algorithm style updating of the population where agents' probabilities of being reproduced are proportional to their respective payoffs in the previous round. Hales finds that in some, but not all, conditions cooperation tends to be quite stable. But the situation is clearly more complex in terms of the longevity of cooperative groups (agents sharing the same tags) that game theoretic formulations would suggest. The Hales paper thus indicates that the sorts of issues addressed within game theory may be better addressed using MABS approaches which are less restrictive and more readily describe real social actors, their interactions and the consequent social processes. Nonetheless, Hales' paper is a long way from representational social simulation as reported by Hemelrijk, Rouchier *et al.* and Downing *et al.* It remains to be seen whether and how Hales' highly abstract model will inform representational models of cooperation and defection in real societies.

We turn to the triad of papers dealing with the role of formal logics in agent based simulation. Teran *et al.* use the elements of SDML that make it consistent with a fragment of strongly grounded autoepistemic logic as a theorem prover for a model implemented in SDML. This is a different approach from that used in (say) Concurrent MetaTem (*e.g.*, [9]) where the program is written in an executable logic. The point of the Teran *et al.* paper is to prove that all *possible* runs of a simulation model will yield outputs conforming to the theorem. This is an alternative to Monte Carlo studies by yielding definite rather than statistical statements of model properties. And it is evidently more convincing than reports of outputs from a few runs of a given simulation model. It also introduces an envelope of what social properties, for example what norms, can emerge from agent interaction within a given model. The David *et al.* paper reports an integrated model with utility-oriented parameters (but not utility maximisation) and a Newell-Simon approach to the representation of cognition to simulate coalition formation based on mutual dependence as an alternative to trust. The dynamics of the model are investigated by means of simulation experiments as the basis for their analysis of appropriate specifications of agent rationality for applications to large, poorly understood (*i.e.*, messy) systems.

Less directly formal, but dealing with extensions to the use of BDI formalisms, is the paper by Norling *et al.* Their purpose is to extend BDI-inspired languages by augmenting the representation of agent cognition to give the agents more human-like decision making strategies. These strategies are derived form the natural decision making (NDM) literature which is specifically geared to agent decision making in messy, uncertain environments with significant agent interaction. This is, of course, very different from the environment – the mechanism and agent designs – recommended by Wooldridge and Jennings for, *inter alia*, BDI agents of the Rao-Georgeff type.

All of the MABS2000 papers make some clear contribution to the development of agent based simulation as a means of analysing large scale, complex systems involving substantial interaction among agents. Many, though by no means all, such systems are social systems populated by humans. Software systems populated by software agents are increasingly important and modelling those systems to inform mechanism and agent design is an important application of the modelling methodology and technology being developed for agent based *social* simulation. Those MABS-2000 papers that do not incorporate demonstrator models either identified new areas of research or new means of addressing existing areas of research.

The final paper in the volume, by Edmonds, amounts to a review essay on this volume. Edmonds takes the papers (including his own!) as exemplars of how we *actually* do multi agent based simulation in comparison with a set of reasonable criteria of how we *ought* to do it. On balance, Edmonds takes a pragmatic approach: what is it we intend to achieve with MABS and what would be required to achieve it. In general terms, Edmonds' view is consistent with the state of MABS research as reflected in this volume: start with the problem. Everything else is in service of the analysis of that problem.

4 Open Research Questions

This section comprises a report of the open discussion sessions at the MABS-2000 workshop.

4.1 Validation

A key issue that attracted considerable attention was validation. The particular issue addressed by the workshop was whether validation should rest on the goodness of predictions or the goodness of the agent and mechanism designs and resulting outputs as descriptions. While there was considerable support for the view that good validation of social simulation models requires prediction, it was also clear that the six papers directly addressing questions and reporting models of real social processes and outcomes appealed to their descriptiveness for the validity of their models. Axtell's simulation analysis of changing retirement ages started from the observation of long lags between a reduction in the age at which Social Security pensions could be drawn and the convergence of actual retirement ages to the pensionable retirement age. The Archisim traffic simulation model was validated by its coherence with actual driver behaviour as determined in psychological studies as well as with actual traffic patterns. Downing, Moss and Pahl-Wostl emphasised the importance of descriptiveness in validating models – partly since no one in their right minds would seriously seek to predict climate change and social responses to the physical and biological effects. Rouchier *et al.* demonstrated that observed heterogeneity among actors must be replicated in software agents representing those actors if observed social phenomena are also to be replicated by the simulations. This amounts to the multi-level descriptive validity sought by Downing *et. al*. Similarly, Hemelrijk was justifying her alternative explanations of inter-sexual dominance among primates by specifying plausible agent behaviour that yields observed social interactions. Schillo, Fischer and Klein validated their analysis of micro-macro correspondences by appeal to the descriptive accuracy of their simulation of real shipping and freight arrangements.

No paper in MABS-2000 sought to predict what has not already been observed. One explanation of this tendency is the natural adoption of real systems as simulation targets and the consequent requirement to devise agents that achieve known ends. This tendency is what ensures that multi agent simulation remains close to that part of the agents research community which starts with well defined problems that are external to their techniques.

4.2 Micro-Macro Issues

The relationship between agent design (the micro) and system performance (the macro) was something of a *leitmotif* in the workshop and discussions. There was a clear argument for (and no disagreement with) the proposition in Sawyer's paper that emergence is a (perhaps *the*) key link between the micro and the macro behaviours and that emergence requires complex interaction among a substantial number of agents. However, no one defined either "complex" or "substantial" in this context. So one important and open research issue remains the relationship between different specifications or classes of agent interaction, system scale and emergence or the consequent relationship between individual and system behaviours.

We know from two of the papers, Davidsson and Rouchier *et al.* how sensitive emergent outcomes can be to agent specifications – a result also developed by Parunak [16] in the previous MABS workshop. However, in the discussion, Axtell pointed to a universality phenomenon of some importance to modellers associated with the Santa Fe Institute (SFI). Universality refers to the emergence of the same macro behaviour from a variety of agent specifications. Two views were enunciated here. One was Axtell's to the effect that universality makes agent specifications less important than would be suggested by several of the papers and methodological discussions (*e.g.* Sawyer, Downing *et al.*, Rouchier *et al.*, Davidsson). The alternative view, expressed by Drogoul, is that participatory modelling approaches – involving drivers in traffic simulations or stakeholders in models developed with the participation of stakeholders, for example – gives central importance to agent specifications. This would seem to bring us back to the validation issue. It may also define the difference between the approach to agent based simulation modelling identified with the SFI and that identified with the European special interest group on Agent Based Social Simulation (ABSS). The SFI approach is concerned with validation as prediction while the ABSS approach is concerned with validation as description. This difference coheres with the largely European concern as exemplified by Castelfranchi, Conte and their colleagues with a foundational approach intended to clarify agent concepts to describe phenomena such as trust and belief associated with individuals.

4.3 Verification

Foundational approaches to agent based simulation are – because of their formalism – frequently presumed to be associated with verification rather than validation. The foundational papers in MABS-2000 – especially Pedone and Conte, Schillo *et al.* and Sawyer – though about foundational issues were directly addressed to supporting the validation of agents as representations of humans. This would actually seem to be the normal role of foundational approaches in agent based social simulation and therefore in software for messy systems that draw on human societies for the analogies used in agent and mechanism design of applications software for messy systems.

5 Some Concluding Remarks on Social Theory

The MABS community has extraordinarily broad interests stemming from the use of simulation models

- to restate and assess extant (usually sociological and anthropological) theories
- the use of sociological theories and concepts to inform simulation models or, more abstractly, simulation modelling
- simulation models as a language of description (so a kind of formalised history and journalism)
- simulation models for scenario analysis,
- simulation modelling as a process of policy analysis and formation.

Orthogonal to these interests are concerns to develop general theories and approaches. The purpose of foundational agent based social simulation is avowedly to develop a general social theory. Representational social simulation by and large ignores theoretical questions. There have been attempts to consider how general individual models might be by aligning several of them [1] or restating them in a more general framework (me). But this is a long way from building a general theory to inform social simulation.

It is surely fair to say that there is no universally accepted general theory of social process. In relation to MABS, the role of a general (or any) theory could only be to constrain (in the terminology of Edmonds) model specifications. The constraints will naturally depend on the purpose of the model. Validation, as Edmonds points out, is one source of constraint. But the point of a general theory is presumably to constrain models by verification – by demonstrating that the models are consistent with some formal theoretical or logical structure. This was clearly the purpose of the sociological contributions to this volume by Sawyer, by Schillo *et al.* and by Pedone and Conte.

An important question for the future development of MABS is whether and how theoretical structures should be used to constrain our models. If sociological theory is to constrain our models, should it be a social realist or a methodologically individualist theory or some synthesis of the two? This question has been asked trenchantly on a number of occasions by, for example, Conte and Castlefranchi [7] and Gilbert [11] and, of course, Giddens [10]. The constraints implied by these different approaches could hardly be more different. This is enough to make representational social simulation modellers extremely wary of the adoption of any extant social theory.

Of course, it would be lunatic to suggest that, without any theorising, model specification is unconstrained *a priori*. Every modeller has a set of techniques and approaches that he or she uses in building models for particular purposes. My own favourites are a development of Paul Cohen's [6] endorsements scheme and, for the more complicated representations of cognition, the problem space architecture of Soar and ACT-R. I find that these generally enable me to capture with my simulation models the behaviours of a wide range of social actors and institutions ranging from domestic consumers of water to formerly state owned enterprises in the Russian Federation. Other modellers use genetic programming or genetic algorithms or simulated annealing algorithms or a host of other techniques. In each case, we are constraining our models by our own predilections based on habit and experience.

There would be clear advantages in having a set of constraints that consistently supported model specifications that were always validated in known, specific conditions. If we had such constraints, it would surely give us more confidence in applying models specified within those constraints to the analysis of social policy and mechanism design for messy software systems. Since we cannot produce an unlimited number of models to generate an unlimited number of scenarios for ill-defined and poorly understood social, natural or software systems, it is clear that some constraints will be essential. A large and growing body of well validated models with shared constraints would surely enhance our confidence in the goodness of similarly constrained models in conditions where the possibilities for validation are limited by the messiness of the target systems.

It may be useful to consider these issues in relation to validation and verification.

Validation is obviously an *a postiori* constraint while verification is an *a priori* constraint. That is, when our models produce output (system or agent behaviours) that are not observed in the target systems (*i.e.*, the real or software systems our models represent), then the models can be respecified to bring the simulation outputs into line with observation. Verification of models with respect to some logic or theory limits the specification of the model before it is used to generate any output at all. So a failure of validation of a model verified against some independent theory or formalism implies either that the model is to be modified (hence divorced in some way from the theory/formalism) or the theory/formalism is modified to support some necessary modification to the model. In the second case, the theory becomes better grounded in observation.

Acknowledgements

I had many useful comments on the previous draft of this introduction from Olivier Barrateau, Bruce Edmonds, Barbara Grosz, Keith Sawyer, Jaime Sichman, and Gerhard Weiss. All of the advice was good and most of it was accepted. I am grateful to them all for the time and trouble they took. Any remaining errors, omissions and infelicities are of course my responsibility.

References

1. Axtell, R, R. Axelrod, J. M. Epstein and Cohen, M. D. (1996) Aligning Simulation Models: A Case Study and Results. *Computational and Mathematical Organization Theory*, 1: 123-141.
2. Brainov, S. and T. Sandholm, "Reasoning About Others: Representing and Processing Infinite Belief Hierarchies" in *Proceedings of the Fourth International Conference on Multi Agent Systems*, (ICMAS-2000), (IEEE Computer Society), 71-78.
3. Bussmann, S and K. Schild (2000), "Self-Organizing Manufacturing Control: An Industrial Application of Agent Technology" in *Proceedings of the Fourth International Conference on Multi Agent Systems*, (ICMAS-2000), (IEEE Computer Society), 87-94.
4. Castelfranchi, C. and R. Falcone (1998), "Principles of Trust for MAS: Cognitive anatomy, Social Importance and Quantification" in Y. Demazeau, *Proceedings of the Third International Conference on Multi Agent Systems*, (ICMAS98).
5. Chandler, A.D. Jnr (1961), *Strategy and Structure* (Cambridge MA: MIT Press).
6. Cohen, P.R. (1985), Heuristic Reasoning: An Artificial Intelligence Approach, (Boston: Pitman Advanced Publishing Program).
7. Conte, R. and C. Castelfranchi (1996), "Simulating Multi Agent Interdependencies: A Two-Way Approach to the Micro-Macro Link" in K. Troitzsch, Y, Mueller, N. Gilbert and J. Doran, Social Science Microsimulation (Berlin: Springer Verlag).
8. Edmonds, Bruce (1998) , "Social Embeddedness and Agent Development", *Centre for Policy Modelling Technical Report 98-46* http://www.cpm.mmu.ac.uk/cpmrep46.html.
9. Fisher, Michael (1997), "A Normal Form for Temporal Logic and its Application in Theorem-Proving and Execution", *Journal of Logic and Computation* 7(4): 429-456.
10. Giddens, Anthony (1984), *the Constitution of Society* (Cambridge: Polity Press).
11. Gilbert (1995)
12. Granovetter, Mark (1985), "Economic Action and Social Structure: The Problem of Embeddedness", A*merican Journal of Sociology* 91(3): 481-510
13. Moss, Scott (1999), "Relevance, Realism and Rigour: A Third Way for Social and Economic Research", *Centre for Policy Modelling Technical Report 99-56* http://www.cpm.mmu.ac.uk/cpmrep56.html.

14. Moss, S. (2000), “Applications-Centred Multi Agent Systems Design (with Special Reference to Markets and Rational Agency)” in *Proceedings of the Fourth International Conference on Multi Agent Systems*, (ICMAS-2000), (IEEE Computer Society), 199-206.
15. Nwana, H.S. and D.T. Ndumu (1999), “A perspective on software agents research”, *Knowledge Engineering Review* 14: (2) pp. 125-142
16. Parunak H.V.D., R. Savit and R.L. Riolo (1998), “Agent-based Modelling versus Equation Based Modelling: A Case Study and Users’ Guide”, in J.S. Sichman, R. Conte and G.N. Gilbert, *Multi Agent Systems and Agent-Based Simulation*, Lecture Notes in Computer Science 1534: 10-26.
17. Penrose, E.T. (1959), *On the Theory of the Growth of the Firm*, (Oxford: Basil Blackwell).
18. Rosenberg, N. (1983), *Inside the Black Box* (Cambridge: Cambridge University Press).
19. Sullivan, D., B. Grosz and S. Kraus (2000), “Intention Reconciliation by Collaborative Agents” in *Proceedings of the Fourth International Conference on Multi Agent Systems*, (ICMAS-2000), (IEEE Computer Society), 293-300.
20. Wooldridge, M. and N. Jennings (1998), “Pitfalls of Agent-Oriented Development” in K. P. Sycara and M. Wooldridge, *Agents '98: Proceedings of the Second International Conference on Autonomous Agents*, ACM Press,

The Use of Models - Making MABS More Informative

Bruce Edmonds

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

Abstract. The use of MABS (Multi-Agent Based Simulations) is analysed as the modelling of distributed (usually social) systems using MAS (Multi-Agent Systems) as the model structure. It is argued that rarely is direct modelling of target systems attempted but rather an abstraction of the target systems is modelled and insights gained about the abstraction then applied back to the target systems. The MABS modelling process is divided into six steps: abstraction, design, inference, analysis, interpretation and application. Some types of MABS papers are characterised in terms of the steps they focus on and some criteria for good MABS formulated in terms of the soundness with which the steps are established. Finally some practical proposals that might improve the informativeness of the field are suggested.

1 Introduction

What is MABS? It could be:

- *Entertainment* – a sort of intellectual computer game where one sets up an artificial system with lots of agents and then play with it to see what sort of effects one can get;
- *Art* – MAS designed and/or constructed for others to admire and enjoy;
- *Illustration* – multi-agent systems designed to animate or otherwise illustrate some sociological, philosophical or mathematical principle, in other words, a sophisticated pedagogic tool;
- *Mathematics* – using simulation as a stand-in for symbolic deduction in distributed systems where such deduction is impractical;
- *Communication* – multi-agent systems as an interactive medium for social exploration, negotiation and communication; or
- *Science* – multi-agent systems as a tool for understanding observed systems.

All of the above are legitimate uses of multi-agent systems. Each has different goals. Each has different roles in society. Each has different criteria for success. It is not obvious that academics who attend MABS workshops have decided what MABS

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 15-32, 2000.

is. Indeed, it is not immediately obvious that there is a need to decide – these different activities can have much to contribute to each other.

However, it is likely that if these different activities are conflated in a single paper then only confusion will result. These activities have very different aims, and so it is unlikely that they can be satisfied simultaneously. Further if the aims of a piece of work are unclear there is a considerable danger that it will be misinterpreted – for example, if a MAS was designed as an illustration of some philosophical principle then it would be a mistake to take this as an indication of the behaviour of any particular natural system.

In *this* paper I will only examine the use of MAS as a tool for understanding – i.e. MABS as a science. In doing this I well understand that these other sorts of activities will be involved, but that the *ultimate goal* of the methodology I will be examining is to gain some understanding by people of systems that are observed. This does not mean that there will not be meaningful *sub-goals* (as will become clear) or that sometimes there will also be *other goals*[1] for any particular piece of academic work (e.g. when a model is to be developed *with* some stakeholders).

To do this I will introduce an abstraction of the scientific modelling process and use this as a *framework* for analysing the activity of MABS as a tool for understanding observed systems. I argue that this analysis is appropriate, does indeed characterise some of the papers in this volume, enables clearer judgments about their precise role and provides a basis upon which to judge their success. I conclude by applying this analysis in terms of practical suggestions to make MABS more informative about observed systems and, therefore, a better science.

2 An Analysis of Modelling

Several philosophers (e.g. [16, 17, 25]), who have observed the process of science have delineated the same picture of modelling. They see a model as enabling an inference process such that it corresponds to some aspect of an observed process. That is, if the inference process is set up so that its initial conditions are appropriate to observed system's initial state then the results of the inferential process will predict the relevant aspect of a later state of the observed system (to some degree of accuracy). This is illustrated in figure 1 below.

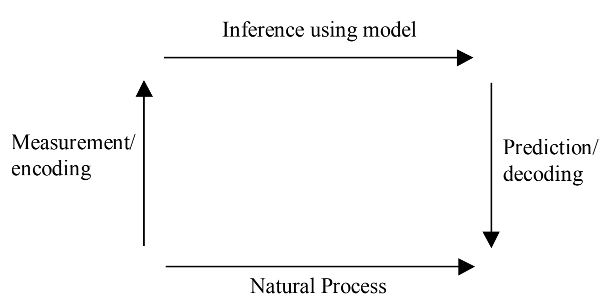


Fig. 1. The basic modelling relation

1 I do not want to imply that these other goals are less important.

In this picture something *is* a model of something else if the above diagram *commutes*, that is to say that if you follow the two different routes around the rectangle then you get the same result.

Frequently in the natural sciences, the model has been encoded as a set of mathematical equations with numbers used as the basic currency for initial conditions and predictions. It was (and is) the job of mathematicians to discover and check the inferential machinery and the job of scientists to discover and check the mappings to and from the mathematical model.

In an imperative computational simulation the model is encoded as a program and the inference is performed by executing that program. In a declarative computational simulation the model is encoded as a set of logical statements and relations and the inference is done by an inference engine acting on these statements. In either case, it is the job of computer scientist to design and check the simulation process.

3 Understanding Multi-Actor Systems by Modelling with MAS

Multi Agent Based Simulations attempt to model a multi actor system with a multi agent system. The modelled actors can be almost anything including: humans, institutions, agents, robots, programs, computers, objects, concepts, positions or even "balls" of water. Whatever the nature of the objects in the target system, the aim is the same: to understand the working of that system through the construction of MAS models and the analysis of their behaviour when run.

The basic sequence is this:

1. a target multi-agent process is chosen and an MAS is designed so as to incorporate the relevant aspects of the target system's structure (*design*);
2. the MAS is then run (this is basically a form of *inference*);
3. the resulting process is analysed by a variety of means, which can include: simple inspection, Monte Carlo, visualisation techniques, or statistics (*analysis)*;
4. finally, the *point* of the whole exercise is to conclude something about the target system, by interpreting any conclusions about the behaviour displayed in the MAS runs back in terms of that system. This interpretation can be as strong as a concrete numeric prediction about the behaviour of the target system or as weak as an indication of possible behaviours expressed in qualitative terms (*interpretation*).

This basic sequence holds for almost any formal modelling and MABS is no exception to this. Of course, the above characterisation of the MABS modelling process is simplified, in part because only rarely is direct modelling attempted. Typically several other layers of models are involved, which are not always made explicit. For example, in many papers in this volume, it is not the target system that is modelled but rather an abstraction of that system which is related to the target system in a vaguer way. Thus we have the picture in figure 2. The MAS is a formal model of the abstraction and the abstraction is an analogical model of the target system (or a class of target systems).

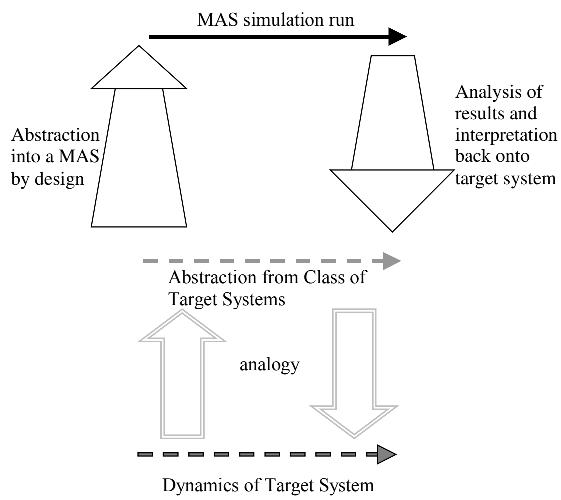


Fig. 2. Modelling using MAS and an intermediate abstraction

In such a case there will be more stages to go through before we are able to infer something useful about the target system. First we abstract from the target system (*abstraction*); then we use MAS simulations to improve our understanding of our abstraction (*design*, *inference, analysis* and *interpretation*); and finally we apply our knowledge of the abstract system to the target system (*application*). This sort of "modelling at one remove" can be a very effective aid to decision making: the MAS modelling is used to hone one's intuitions and this increased understanding is implicitly utilised when making decisions with regard to the target system. This is what Moss, et al. [24] discovered occurred in the development and use of UK macroeconomic models – the direct predictions of the models were almost completely useless on their own, but could be helpful to the experts that used them as an aid for thinking about the issues. The most common error in this sort of modelling is that modellers conflate the abstraction and the target in their minds and attempt to interpret the results of their MAS model directly in terms of the target system. This can manifest itself in terms of overly strong conclusions in MABS papers, for example in [1] at the end of chapter 5 Axelrod says: *"… Darwin's emphasis on individual advancement has been formalized in terms of game theory. This formulation establishes conditions under which cooperation based on reciprocity can evolve …"*.

The determination of the relevant abstraction can itself be analysed into two parts: deciding on the language/theoretical framework in which the abstraction will be specified; and formulating the abstraction within that framework.

The human cost of searching amongst candidate formal frameworks/languages is sufficiently great that one person usually has effective access to only a limited number of such frameworks. For this reason the framework is rarely changed once it has been chosen, rather people will go to considerable lengths to reformulate within a framework rather than change it. Added to this, the framework affects one's perceptions of the problem itself so that people 'see' problems in terms of their favourite framework [18]. Both of these effects mean that instead of the domain (i.e.

the target systems) determining the most appropriate framework, the formulation is adjusted so that the target systems can be mapped into the framework even if this means distorting the target system almost beyond recognition.

4 Attempting to Improve Insight

Given the basic situation as described above, the question arises as to how one can improve the insight into the target systems in such modelling both in terms of the relevance and reliability of any conclusions. One way of approaching such a task is to improve any of the individual steps involved: abstraction, design, inference, analysis, interpretation or application.

4.1 By Increasing Strength of Formal Modelling Steps

The 'hard' sciences have often been characterised as concentrating on strengthening the formal modelling steps, and perhaps the most developed of these is the inference step. Utilising formal modelling techniques (including computational, numerical and symbolic models) ensures the reliability and consistency of this step. In conjunction with the rest of the formal modelling steps (design, analysis and interpretation), this forms the core of the modern "scientific method".

The danger of this approach is that improving the reliability of the formal modelling may be at the expense of its relevance to the target systems. The strength of the formal steps may be increased but the overall chain has been weakened because this is more than offset by the increased irrelevance of the abstraction to the target systems. In other words, the overall chain has been weakened.

A hallmark of many 'degenerate research programs' [19] is that they concentrate on the purely formal aspects at the expense of their relevance to the target systems in their domain. It seems to be that a lack of empirical or practical success motivates a retreat into pure formalism. Particularly unsuccessful fields may retreat so far that the field concentrates upon a single step: the inference step. Such fields are marked by glut of purely formal papers and total irrelevance to any real systems – they lack the generality of pure mathematics *and* lack the relevance of good science.

4.2 By Improving the Correspondence with the Target System

Another well-established approach is to improve the analogical relations between the target systems and the abstraction, in other words to focus upon the descriptive aspects of modelling. Typically such an approach is marked by the presence of much richer and more complex abstractions, which lack the 'purity' of abstractions found in the 'hard' sciences. This approach is archetypal of the 'softer' social sciences[2].

This approach has the opposite pitfalls: it may be that the inference steps are so vague so as to be completely unreliable. This unreliability can mean that the

[2] Although not always, [31] is a call to eliminate the abstraction-target system gap in agent-based computer languages, since for the most part the connection between formal systems (such as BDI logics) and the computer architectures (e.g. Mars) they where supposed to underpin is suggestive rather than strict.

abstractions do not even have sufficient coherency to allow a consensus about the meaning of parts of the abstraction in terms of the target systems to be developed. This can result in a situation where the academics involved are all using such terms in subtly different ways and discussions give the impression of "ships passing in the night". Again, it is the strength and relevance of the complete chain that counts.

It is notable that the natural sciences spend a huge amount of time and effort in establishing the correspondence of their abstract and target systems. This is not often apparent in a causal inspection of the output of these sciences, but is covered by all those involved in applied science. Indeed the activities of developing new measurement techniques, experimentally testing theory and applying theory in technology accounts for the greater part of the effort in such sciences. Thus although it has been a common perception that the hard sciences are marked out by their use of formal models, a deeper characterisation might point to the relatively developed and precise (though rarely simple [5]) correspondences between their abstractions and their target systems. It is this correspondence which is the engine of their success.

There are two ways to achieve greater descriptive correspondence: to choose an appropriate but specific framework; or to use a very expressive framework which embodies few and relatively weak assumptions about the subject matter. The first requires effective access to a relatively large number of frameworks and the conditions of application of each framework to be known. The second requires great expertise in the powerful framework so that a useful model can be formulated in it – this is a problem because the more powerful the framework the more difficult it is to calculate or prove anything. Thus the adoption of a very expressive framework can have negative effects such as shifting the difficulty to the design stage where one has to get a simulation to run, or again distorting the target system so that a relatively *simple* model can be fixed upon[3].

5 MAS as a Modelling Paradigm

What distinguishes MABS from other modelling enterprises is its use of MAS for the formal models. This represents a small step towards descriptive realism because it implies a commitment to analyse the target system in terms of the natural system boundaries that exist there. In other words it is almost universal in MABS to map objects, actors or other natural entities in the target system onto agents in the MAS, so that the 'boundary' of the entities correspond to those of the agents and that the interactions between entities correspond to interactions between agents. This is in contrast to when some agents are represented via average properties or single 'representative agents' [4]. As in any modelling enterprise, the choice of MAS as the modelling paradigm will have consequences in terms of the reliability and relevance of the trade-offs described above. It does not, of course, avoid these trade-offs, but does allow a new trade-off to be exploited and has the potential to push the boundaries of both reliability and relevance forward. In the subsections below I summarise some of the consequences of adopting MAS as a modelling framework.

---

[3] Even a sufficiently expressive formal system can bias ones modeling as inevitably some sorts of model in that system will be simpler than others [10].

5.1 MAS as a Step towards Greater Descriptive Accuracy and Precision

The use of MAS allows the application of formal models (computational models) to social systems (in its widest sense of interacting distinct entities) without the loss of relevance which has accompanied some attempts at such formal modelling (e.g. [14]). The MAS can be constructed so that the agents in the model and their interactions *do* correspond to those observed in the target systems. Thus the processes of abstraction and application become easier and more transparent.

Of course, MAS can easily be used so they do not improve either the reliability or the relevance of more traditional modelling techniques. The agents may not correspond to anything in the target systems, as happens in many evolutionary models where there is only a vague population to population relation [6]. Another possibility is that the 'agents' in the model are *so* abstract that the relation between them and real target systems is, at best, merely suggestive.

5.2 More Difficult to Infer General Results

One consequence of the move to increased descriptive accuracy that is implicit in the use of MAS to model systems is the consequent particularity of any model outcomes. This lack of generality comes about in, at least, two ways. *Firstly*, the ability to build MAS models so that there is considerable structural correspondence between it and the target systems brings to the fore the generality-relevance trade-off. If the abstraction closely corresponds to a small class of target systems, then the conclusions can only be safely applied to those systems, and if the abstraction is more distant, that is, it attempts to capture some common aspects of a larger class of targets then the relevance of the abstraction to those systems becomes harder to establish and any conclusions less reliable. *Secondly*, there are inevitably some processes in the target systems (e.g. internal to the target entities) that are not explicitly included in the model (either they are unknown, deemed irrelevant or impractical to implement[4]). It is a common tactic to substitute an indeterministic element into the MAS model to stand instead of this process, usually in terms of a random number or choice. In such a case it is not expected that an individual MAS trajectory will correspond to that of the abstraction (and hence of the target systems) but that the *collection* of trajectories corresponds to the possible trajectories of the abstraction.

5.3 Greater Contingency in Inference

The fact that most MABS are not deterministic, for the reasons noted above, means that the inference represented by the running of the model is modulus the indeterminism in that simulation. In simple non-agent-based simulations it occasionally happens that the randomness cancels out when a suitably large number of runs are performed. In these cases a Monte Carlo approach can be used and the range and central tendency of simulation behaviour safely deduced. In almost any agent-based simulation (and almost all simulations I have actually come across) one can not safely assume that random elements introduced into the structure will be uniformly carried through into the outcomes, so that one can safely generalise about

[4] These three reasons are frequently conflated in reports of simulations and only rarely explicitly distinguished and documented.

'central tendencies', equilibria and the like[5]. This contingency means both that single runs of the simulation may be completely unrepresentative of the system's general behaviour (if indeed there is such a 'general behaviour'), and that performing many runs of a simulation and averaging the results may merely result in a misleading artefact. In such cases a painstaking examination of single runs is often necessary in order to distinguish what is happening in each so that one can begin to determine how to classify the simulation trajectories. The increased descriptive realism has meant that the simulation has imported more of the object system's behaviour, including its unpredictability and complexity.

5.4 Syntactic Complexity Can Imply a Different Language for Results as for Specification

The complexity of the interactions and internal processes of many MABS means that it is often impractical to trace the trains of causation backward to determine a small set of causes for observed behaviour. Rather, as one traces the computation back the formal causes multiply until it encompasses almost the whole computation and every agent in the system. This is an example of 'causal spread' [29] and is common in complex and distributed systems. Sometimes it is the case that a *better*[6] explanation can be made in terms other than that of the systems detailed computation (e.g. an increase in 'entropy'). This is a clear example of *emergent* phenomena: the phenomena are not easily explained from the specification and detailed computation of the system, but make sense only within a *new* framework for their representation. The practical import of this is that the analysis and interpretation stages of the modelling process require much more attention than in simple deterministic or stochastic mathematical models.

5.5 Greater Variety of Possible Models

The move to a more descriptive modelling stance that is implied by the use of multi-agent models, and the effective encapsulation of the agents means that there are many more models that are specified differently (for example with different agent learning or reasoning mechanisms) but have essentially the same results (from the point of view of the intended interpretation). In such a case there is no single 'correct' model, but a whole class of models that give adequate results. It has been traditional to choose the 'simplest' adequate model as the 'correct' one, but there is no reason to suppose that such a model would be more representative of the target system's behaviour when used in different domains. Such 'simplest' models have pragmatic value in terms of ease of use and didactic value but are not justified on the grounds of being a better (or more likely) indicator of the truth [13].

---

[5] Interestingly this is true even of deterministic agent-based systems, because the complexity of the system means that many aspects of the simulation are practically unanalysable, except as effectively indeterministic elements. For example, this is true of a pseudo-random number generator – which is precisely its point.

[6] Of course, what constitutes 'better' is the core of a large philosophical dispute, which I do not have space to go into here.

The practical upshot of this is that it is desirable to constrain the space of candidate models as much as possible. In particular there is a *greater* need for such constraints in MABS than in simpler types of models where there are fewer possible variations. This is why the processes of verification and validation are particularly important to MABS: verification enforces the constraint of its intended *design* (i.e. its conformance to the abstraction of the target system); and the validation ensures that the processes that result from running the models are acceptable in terms of outcomes [23].

Strengthening verification requires that the abstraction (on which the design of a simulation is based) is specified as unambiguously as possible, and that the actual implementation of the simulation can be checked against this specification. The weakest possible verification is when the abstraction is not described directly but is only implicitly indicated via descriptions of the target systems and the actual simulation. In such cases it is impossible to tell which features of a simulation are intentional and which are merely necessary in order to get it to run [7].

Strengthening validation means checking the output of the simulation in as many ways as possible by comparison with the system goals or actual target system behaviour. The strongest possible validation involves checking hard output at all stages and levels of detail of the model against unknown data gained directly from target systems to within pre-declared limits. The weakest validation is where the simulation is merely claimed to exhibit some qualitatively described behaviour. Validation which merely checks that the simulation behaves as intended does not add any additional constraints than those already introduced by verification, because it is essentially a design check.

6 Some Archetypes of MABS Papers

To illustrate the analysis above, I present some archetypes of the sort of papers one finds at MABS conferences and journals followed by an analysis of an example of each taken from this volume[7]. These are in the order of the modeling steps that they focus on. They are also, in my opinion, roughly in order of increasing usefulness and rarity. Unfortunately the former have the academic status because they are more often cited in reports of the latter even if they have, in fact, contributed little to their success. As a generalization the field spends far too little effort on the later stages of the modelling process (analysis, interpretation and application).

6.1 Establishing an Abstraction

This sort of paper seeks to establish that a certain way of framing/describing/ formalizing aspects of the target systems is necessary. The abstraction can be justified on the basis of either: *a priori* reasoning (their own or others) or on case studies of the target systems. Good examples of this sort of paper give indications of *how* they might be modelled in a simulation and *when* the proposed abstraction is appropriate. This sort of paper is only finally justified when the abstraction it

[7] To be precise the examples are taken from the MABS2000 pre-proceedings, as I did not have access to the final accepted versions. I strongly suspect that my characterisations will still hold in the final versions, but I suggest that you check this by reading them.

proposed leads to simulations that, in their turn are informative about the target systems. Common forms of this include: proposals for abstract architectures and formalisations (e.g. a logical formalisation of trust).

A clear example of this type of paper is by Sawyer [26]. This suggests a framework for modelling collaborative emergence in social systems based upon the thought and observations in sociology over the last century. This is primarily an argument as to what is *required* in order to model collaborative emergence via downward causation, with suggestions as to how such a simulation model might be designed. There is no suggested validation test for the success of such a simulation, but rather is motivated by an example of such collaborative emergence of dialogue from an improvised theater-piece exhibited. A couple of applications for a simulation incorporating this structure are suggested so an implicit validation for this abstraction is whether this does turn out to be a useful way of implementing these applications. The range of systems that this will be useful for is not defined except that it pertains to 'natural social systems' – a category which is very broad. It is thus difficult to build upon the work in this paper because one does not have a clear idea of the assumptions that are necessary in order for the stated conclusions to hold. It would be helpful to give an indication of *when* it is necessary to include the sort of structures described in this paper in a system (unless the implication is that one *always* needs it for any complex social system); and it would also be helpful to be given some way of knowing if one has succeeded. One could attempt to build upon this paper by *guessing* that it might be useful, trying to use it to build a system for one of the suggested purposes and then analyse whether the suggested structure did, in fact, help.

6.2 Documenting a Design Proposal

Here the abstraction is more or less given (often in previous papers by the authors or others) and a simulation design is proposed along with its justification. The merit of this sort of work is where is opens a way to implementing some aspect that was previously thought impractical. Its justification comes when simulations built upon the design come through with the goods, in other words the simulations do capture the aspects claimed for it in a verifiable way, they run with reasonable use of computational and human resources, and give results that under analysis are informative about the behaviour of the abstraction.

An example of this type of paper is by McGeary and Decker [20]. This basically documents the design of a simulation of a Food Court. The abstraction of the food court and the system/agent architectures has already been decided upon and this document brings these two elements together. This abstract design is illustrated by one example: a more specific design for the process by which a Virtual Food Court (VFC) hires a (virtual) waiter. This design is situated by giving a little of the background to their model. Its purpose is to provide a system which people could use to experiment with in order to "*…explain, step-by-step, some specific and particular economic phenomena …* [because we believe it] *is necessary to detail the conditions under which the studied phenomena happen …*" (section 3 second paragraph). The proof of this will be if and when their goal is achieved using this system. It is difficult to see how this paper helps other researchers – there is not enough detail in it to allow others to duplicate it, and it is unclear which of the structural assumptions made derive from observations of real food courts, waiters etc. which derive from other people's theories and which are merely necessary in order to make implementation feasible. In the end, all it seems to do is keep others informed of what they are thinking so that they stay interested until they actually produce systems or get some results.

6.3 Exploring the Behaviour of an Abstraction

Here some aspect of an abstraction's behaviour is explored using simulations. This is a sort of stand-in for formal inference in analytic models. The relationship with target systems is typically only at the suggestive level. The utility of this sort of paper is (as in pure maths) dependent upon the *generality* of the results and knowing their *conditions of application* – if one can recognize another abstraction as sufficiently similar so that one can use the general results (for example as a guide to a simulation's design) this is helpful. If the results are particular to certain set-ups or the necessary structure for the discovered outcomes is unclear then it is almost impossible to *use* the results of such a paper. Sometimes such papers are used to make claims about the behaviour of target systems in a vague and ultimately unjustified way.

An example of this type of paper is by Axtell [3]. All though some of the models described were obviously motivated and described in terms of observed systems, the objective of *this* paper is to demonstrate some general properties of MAS simulations: that: the topology of interaction; the medium of interaction; and the agent activation regime can each substantially change the results. Axtell is, in effect, saying *be careful about your simulation design and implementation you may be inadvertently biasing your results*. Although these results, stated in this way, seem obvious pointing them out with concrete examples is useful because may researchers *have* assumed that they were not relevant factors in their models[8]. This is a very general result, and a salutary lesson to researchers who want to over-interpret their particular results. However, in any implementation one is forced to make *some* such pragmatic decisions and it would be great if some *conditions* under which one had to be careful about such matters could be discovered.

6.4 Suggesting Solutions to Real Problems

Here the modelling cycle is finished in order to suggest some solutions to some real problems. That is conclusions about the behaviour and processes of an abstraction are applied to some target system in order to suggest solutions to some problem concerning them. In order for this to be done with any reliability there has to be a close connection between the abstraction and the target systems. If this connection is based upon a prior theoretical basis then the connection will be only as reliable as the extent of the practical validation of that theory, if it is a descriptive model then the application will only be meaningful in terms of the same descriptive language.

An example of this type is by Hemlrijk [15]. Here a concrete problem is focused on: explaining why in many group-living primates, males allow females access to

---

[8] It is notable that such 'errors' were often made because of an over-enthusiastic readiness to *apply* results from the analysis of a set of simulations straight to the target systems without regard from the difficulties of detangling the assumptions made in the abstraction, design and implementation of their simulations.

resources exactly when they are sexually attractive. An abstraction of the relevant primate behaviour is described, then a MAS model produced, explored and analysed. The conclusion of this is that there is a possible new explanation of this behaviour, namely that results from male attraction to females and not as a strategy to gain access to females. The model also makes some other predictions in terms of the levels of aggression which seem to match observations of such social systems. This paper is grounded in observations of real systems and completes the loop by applying its results back to those systems in a credible and restrained way. In addition to this the process of model building and the assumptions it makes are fairly clear. This work can be built upon by building alternative models with the same or different assumptions and experimentally or observationally testing the predictions it makes.

6.5 Methodological Papers

A final type of paper is the one that discussed the methodology of MABS itself. This is in a slightly different position to those archetypes above because instead of doing MABS it is talking about *how to do* MABS. In fact, since MABS is still a young area of study, it is common for most papers to make *some* comment upon the methodology. The only justification of such papers is if it helps other researchers think about what they are doing, and do it better. Any amount of theoretical pronouncements about methodology is worthless compared to a single practical suggestion that is actually helpful.

An example of this type is by Edmonds [11]. It does not attempt any formal modeling of its subject matter but remains at the level of an abstraction[9].

7 Towards Criteria for Informative MABS

Given the characteristics of MABS as discussed above, the question arises as to how are we to distinguish the projects that will usefully inform us about the target systems that concern us. In other words, what are the criteria for good MABS work? We are not involved in *engineering* MAS to meet defined goals, so such considerations as the computational complexity of such design [30] are not directly relevant.

The ultimate criterion for any modelling enterprise is whether it *works* – in other words: does it help us to do things, predict outcomes in target systems, build systems that perform as we want etc. This is the *only final* justification of scientific intellectual work – even pure mathematics is justified by the fact that a lot of it *has* turned to be useful to other disciplines. Of course, it is notoriously difficult to predict which work will turn out to be useful, because many pieces of work are only useful when combined with other work. For example, one paper might specify a certain type of model and determine its behaviour and another might establish an application of this model to a target system (the traditional *theoretical* vs. *applied* distinction). The construction of knowledge about complex systems is necessarily a piecemeal and socially distributed project.

What we can do is to try to ensure that the complete modelling cycle is as strong as possible. This involves at least two areas. *Firstly*, that all modelling steps attempted

[9] Although the beginnings of a formal model can be distinguished in [22].

are as sound as possible and as well documented as possible. This maximises the chance that one piece of work can be used by another, because the scope of its applicability will be identifiable and within this scope it is likely to be reliable. *Secondly*, that all the individual projects *do in fact* join up into a complete model of some target systems, i.e. that *all* the modelling steps are completed, are sound and connect up to form a complete chain.

Thus we can devise six process criteria that can help us judge work, one for each of the modelling steps.

1. **Abstraction**. *Is the abstraction specified?* Has it been made clear which aspects of the target system the abstraction is supposed to represent and over what class of target systems is it intended that the abstraction will cover? Is it clear that the abstraction corresponds sufficiently to the target system that it remains relevant?
2. **Design**. *Is it clear how the design relates to the abstraction?* Is it clear which parts of the simulation specification derive from the abstraction and which are details necessitated by the implementation? Is it possible to verify that the design does correspond to the abstraction and that the implementation does meet the requirements of the design? How do we know that there are no critical bugs in the simulation? How do we know that the important outcomes are not critically dependent upon the implementation language?
3. **Inference**. *Is the inference of outcomes sound?* To what extent are the outcomes described a necessary result of the model specification and design? Is the underlying inference process clearly specified and understood? Are the described outcomes representative of all the outcomes and, if not, what other types of outcome are there? Are the outcomes critically dependent upon particular parameter settings? Is it clear where indeterministic elements have included?
4. **Analysis**. *Is the analysis clear?* Has the analysis been well motivated in terms of what it abstracts from the outcomes? Is the analysis replicable from the description? Is the raw data accessible anywhere? Is the analysis technique demonstrably applicable to the outcomes? How reliable is the technique?
5. **Interpretation**. *Is the interpretation justified and relevant?* Do the results of the analysis justify the conclusions? Is the interpretation done into the same framework as the abstraction's specification?
6. **Application**. *Are the conclusions in terms of the target systems justified?* Do the strength of the other modelling steps provide the justification for the conclusions? Does the strength of the correspondence between abstraction and target system justify the strength of the conclusion?

It would be impractical for all these steps to be covered in every paper. Some papers will concentrate on one thing, e.g. the establishment of a suitable abstraction, and some will concentrate on other aspects, e.g. simulation analysis, etc. Of course, there is an onus to document and justify the steps that *are* covered. If a paper only focuses on some part of the complete modelling enterprise it also has an onus to specify two additional sections: its requirements in terms of the preceding steps that might lead up to it and its scope in terms of any steps that might follow from it. This will provide guidance to those who may be working on the missing sections.

For example, if a paper is focused on establishing an abstraction of some aspect of a class of target systems, then those who hope to use that abstraction in order to build a simulation and then to conclude something about outcomes of such an abstraction, will need to know what in the abstraction is essential and what a by-product of the descriptive process (this is especially important if the specification of the abstraction is formal). Those who hope to use known outcomes of a certain abstraction to draw conclusions about the target systems will need to know under what conditions they can safely do this (and indeed they will need to know what the target systems are).

Thus we can enumerate two more criteria.

7. **Conditions of Applicability**. *Are the conditions necessary for the described steps to be applied clear?* Are these conditions even known? Are the conditions directly specified or only implicitly indicated? Would a third person who read the paper know when they could use the work described?
8. **Generality of Conclusions**. *Is it clear in which circumstances the conclusions of the work hold?* Are all assumptions revealed and adequately documented? Is the reliability of the conclusions specified? Would a third person who read the paper know to when and to what the conclusions could be safely applied?

8 Conclusion: Some Ways Forward

With a view to improving the extent to which MABS can satisfy these criteria, I now conclude by making a number of practical suggestions. These are in addition to the obvious one of establishing a norm that published MABS papers meet the above criteria.

8.1 Strengthening Design Methodology

Despite the fact that many MABS papers concern themselves with the abstraction and design stages, the methodology in these areas is weak. This is contrast to the relatively large amount of attention that these steps have received in the wider multi-agent community. Here the ambition is that the abstraction should be specifiable in a logic with known properties and that an implementation that is based on this abstraction should be formally verifiable to that abstraction. So far this has been a more of a hope than a reality (with a few notable exceptions applied to 'toy' problems). It is unlikely that an abstraction that corresponded sufficiently to most of the target systems of interest to the MABS community would be amenable to such a complete verification, but that does not mean that there are not sensible steps that can be taken to strengthen these processes in our domain.

Firstly, as must be obvious from the comments above, there is a great need to explicitly distinguish and specify: what are the target systems of interest; what is the abstraction we are taking from these (including the framework and any assumptions or relevance judgments this involves); what is our intended design for our simulation; and finally what our implementation of our design is. These will all be, almost certainly, *different* – conflating them will obscure the conditions of applicability of the work described and thus impede their utility to others who might otherwise wish to build upon them.

This sequence of distinctions perhaps points towards a deliberately staged process from descriptions and perceptions of target systems up to an implemented simulation. Given the difficulty of verifying the sequence *in one jump*, the obvious thing is to separate it into sensibly small steps, each of which can be more carefully documented and understood. Some of these steps may be amenable to more formal verification and some will be composed of descriptive argument, but at least the *status* and *nature* of each step will be a lot clearer.

There are two tools that could assist in the atomisation of the abstraction-design-implementation process. *Firstly*, a "Social Simulation Specification Language" could help make the design of a MABS clear. This was suggested at a meeting of the Agent-Based Social Simulation (ABSS) SIG of Agent-Link but little progress has been made towards its realisation. *Secondly*, one could envisage the extension of a constraint-based architecture which allows the early execution of fairly 'bare' specifications of simulations, albeit extremely inefficiently. This 'bare' framework could then be incrementally supplemented by additional constraints to make it gradually more efficient until we get to a fairly standard MABS simulation. The 'bare' framework will be a lot closer to the simulation specification and the incremental process could be checked so that it was apparent that the behaviour of the simulation at each stage was closely related to the previous one. It would also have the benefit of forcing out into the open any necessary 'assumptions' that were *not* part of the design. Some of the benefits of this sort of approach can be seen in some of Jim Doran's work (e.g. [8]) and Oswaldo Téran's [28].

8.2 Constraining Model Possibilities

As noted above, one of the effects of using the MABS paradigm is the explosion of possible simulation forms, with the implication that there will be many more simulations that display the same *outward* behaviour in terms of actions of agents. The upshot of this is that we need *as many* constraints upon our MAS models as possible and what constraints we apply need to be documented and checkable.

The most basic example of this is the *separate* checking of the simulation outcomes against a data model of the target processes, in addition to checks about the simulation structure and behaviour with respect to the design.

We should seek to verify and validate our models at as many levels of detail as possible. An example of this is proposed in [9] in which it is proposed that different levels of a social simulation be separately validated against the relevant level of the target systems and verified against each other. If a simulation is of an entity about which there is some knowledge about its internal computational processes, then these should be applied to the simulation unless a good reason can be presented. Maybe, at some point in the future we will know if and when a simpler, more efficient algorithm can be substituted for a complex cognitive process, but this is not presently the case.

Another way of increasing the constraints upon our models is by validating at a finer temporal level. That is to say that the intermediate stages of the resulting processes in the simulation should be checked against those of the abstraction or, even better, against data from the target systems. Just checking some statistics about the final outcomes is far weaker.

8.3 Strengthening the Generality of the Inference Step in MAS

The use of indeterministic elements in MABS simulations is almost always unavoidable and the systems are complex in fundamental ways. These two facts mean that simulations are highly contingent so that it can not be assumed that individual trajectories are *at all* representative. In addition to this the fact that simulations (as the systems they seek to model) can be highly *divergent* in their behaviour means that statistically established central tendencies and measures of spread can be very misleading. Thus we need some tools and methodologies for dealing with the whole *envelope* of simulation trajectories.

Firstly, the indeterministic elements in a simulation should be clearly tagged to indicate what sort of process they are 'standing-in' for (e.g. an unknown process, an irrelevant process, a computationally expensive process etc.).

Secondly tools are needed to control the indeterminism in a simulation and allow the systematic exploration of the complete *envelope* of the trajectories. This could involve a framework which allows the systematic exploration of possibilities in some sort of constraint-based model search (as suggested in [28]).

Thirdly there could be established mappings between different architectures (MAS simulation, constraint-based search, theorem provers, data visualisation tools etc.) so that simulations could be transformed between them to enable different types of exploration to take place with respect to the same simulation design.

8.4 Structured Archive of MABS/SS Papers

One of the problems with the field is that there is very little work that builds on, compares or repeats other work. Notable exceptions include [2] which attempts to 'align' models in different architectures and the "canonical task environments" of [21]. What could help would be a structured archive of MABS results. If there could be a site where different types of MABS paper were deposited with enough information as to what they referred to being indexed it would be easier to start establishing and mapping out the connections between them. For example as each paper is submitted some extra information could be entered to specify which: problems/target systems; abstractions/formal systems; modelling approaches/techniques; simulation implementations; sets of results; sets of analyses of results; and interpretations it refereed to. Each in each of these categories could be give a unique identifier, with new ones being added by authors as they felt necessary. Thus as well as some records enabling browsers to find connections between papers via their linking to these identifiers, a history and comments could be built up about each identifier representing each problem etc.

8.5 More Descriptive Modelling

Finally, I will end with a call for more low-level, descriptive modelling rather than ambitious abstract, high-level modelling that hopes to explain a lot *before* sufficient field work has been done [12]. Historically 'armchair' theorising in advance of sufficient field work has not been successful [27], since our own preconceptions of how things should be is very strong. Rather we might be more successful if we attempt to produce generalisations in a *post hoc* manner, so that data and descriptions from our target systems can guide our modelling.

References

1. Axelrod, R. (1984) *The Evolution of Cooperation*. New York: Basic Books.
2. Axtell, R., Axelrod, R., Epstein, J. M. and Cohen, M. D. (1996) Aligning Simulation Models: A Case Study and Results. *Computational and Mathematical Organization Theory*, **1**:123-141.
3. Axtell, R (2000) Effects of Interaction Topology and Activatioon Regime in Several Multi Agent Models, this volume.
4. Barro, R. J. and King, R. G. (1984) Time-Seperable Preferences and the Intertemporal-Substution Models of Business Cycles. *Quarterly Journal of Economics*, **99**:817-839.
5. Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford : Clarendon Press.
6. Chattoe, Edmund (1998) Just How (Un)realistic are Evolutionary Algorithms as Representations of Social Processes?, *Journal of Artificial Societies and Social Simulation*, **1**(3), http://www.soc.surrey.ac.uk/JASSS/1/3/2.html.
7. Cooper R., Fox J., Farringdon J.& Shallice T. (1996) A Systematic Methodology for Cognitive Modelling, *Artificial Intelligence*, **85**:3-44.
8. Doran, J. E. (1997) Foreknowledge in Artificial Societies. In R Conte, R Hegselmann, and P Tierna (eds.) *Simulating Social Phenomena*, *Lecture Notes in Economics and Mathematical Systems*, **456**:457-469.
9. Downing, T., Moss, S. and Pahl-Worstl, C. (2000) Integrated Assessment: Prospects for Understanding Climate Policy Using Participatory Agent-Based Social Simulation, this volume.
10. Edmonds, B. (2000) The Purpose and Place of Formal Systems in the Development of Science. CPM Report 00-75, MMU, UK. http://www.cpm.mmu.ac.uk/cpmrep75.html
11. Edmonds, B. (2000) The Use of Models, this volume.
12. Edmonds, B. (in press). Towards a Descriptive Model of Agent Strategy Search, *Computational Economics.* Also at http://www.cpm.mmu.ac.uk/cpmrep54.html
13. Edmonds, B. (in press), Complexity and Scientific Modelling, *Foundations of Science*. Also at http://www.cpm.mmu.ac.uk/cpmrep23.html
14. Fisher, M. (2000) Characterising Simple Negotiation as Distributed Agent-Based Theorem-Proving – A Preliminary Report. *Proc. of the 4th Int. Conf. on MultiAgent Systems*. Los Alimitos: IEEE Computer Society, 127-134.
15. Hemelrijk, C. K. (2000) Sexual Attraction and Inter-secual Dominance amoung Virtual Agents, this volume.
16. Hesse, M. B. (1963) *Models and Analogies in Science*. London: Sheed and Ward.
17. Hughes, R. G. (1997) Models and Representation. *Philosophy of Science*, **64**(proc), S325-S336.
18. Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
19. Lakatos, I. (1983) *The methodology of scientific research programmes*. Cambridge: Cambridge University Press.
20. McGeary, F and Decker, K. (2000) Modelliing a Virtual Food Court using DECAF, this volume.
21. Moss, S (2000) Canonical Tasks, Environments, and Models for Social Simulation, *Computational and Mathematical Organization Theory*, **6**:249-275.

22. Moss, S. and Edmonds, B. (1998) Modelling Economic Learning as Modelling, *Cybernetics and Systems*, **29**:5-37. Also at http://www.cpm.mmu.ac.uk/cpmrep03.html
23. Moss, S., Edmonds, B. and Wallis, S. (1997) Validation and Verification of Computational Models with Multiple Cognitive Agents. CPM Report 97-25, MMU, UK. http://www.cpm.mmu.ac.uk/cpmrep25.html
24. Moss, S., Artis, M. and Ormerod, P. (1994) A Smart Macroeconomic Forecasting System, *The Journal of Forecasting* **13**:299-312.
25. Rosen, R. (1985) *Anticipatory Systems*. New York: Pergamon.
26. Sawyer, R. K. (2000) Simulating Emergence and Downward Causation in Small Groups, this volume.
27. Simon, H.A. (1986). The failure of armchair economics [Interview]. *Challenge*, **29**:18-25.
28. Terán, O., Edmonds, B. and Wallis, S. Mapping the Envelope of Social Simulation Trajectories. This volume. Also at http://www.cpm.mmu.ac.uk/cpmrep72.html
29. Wheeler, M. and Clark, A. (1999). Genic Representation: Reconciling Content and Causal Complexity. *British journal for the Philosophy of Science* 50: 103-135.
30. Wooldridge, M. (2000) The computational complexity of agent design problems. *Proc. of the 4th Int. Conf. on MultiAgent Systems*, Los Alamitos: IEEE Computer Society, 341-348.
31. Wooldridge, M. (2000) Computationally Grounded theories of Agency. *Proc. of the 4th Int. Conf. on MultiAgent Systems*, Los Alamitos: IEEE Computer Society, 13-20.

Effects of Interaction Topology and Activation Regime in Several Multi-Agent Systems

Robert Axtell

Center on Social and Economic Dynamics, The Brookings Institution
1775 Massachusetts Avenue, NW, Washington, D.C. 20036 USA
raxtell@brook.edu

Abstract. The effects of distinct agent interaction and activation structures are compared and contrasted in several multi-agent models of social phenomena. Random graphs and lattices represent two limiting kinds of agent interaction networks studied, with so-called 'small-world' networks being an intermediate form between these two extremes. A model of retirement behavior is studied with each network type, resulting in important differences in key model outputs. Then, in the context of a model of multi-agent firm formation it is demonstrated that the medium of interaction—whether through individual agents or through firms—affects the qualitative character of the results. Finally, alternative agent activation 'schedules' are studied. In particular, two activation modes are compared: (1) all agents being active exactly once each period, and (2) each agent having a random number of activations per period with mean 1. In some models these two regimes produce indistinguishable results at the aggregate level, but in certain cases the differences between them are significant.

1 Introduction[1]

One class of multi-agent systems (MAS) consists of a relatively small number of agents, each of whom has relatively sophisticated behavior (e.g., a rich cognitive model, perhaps for dealing with a complex task environment [18]). A different type of MAS involves relatively large numbers of behaviorally simple agents. This second family of multi-agent systems is of significant interest as the basis for empirically-relevant models of human social and economic phenomena. Such models typically involve the use of aggregate social or economic data to estimate parameters of a MAS in which agents have heterogeneous internal states (e.g., preferences) but a common repertoire of behaviors (e.g., economic exchange).

One reason for the elevated attention given to simple agents is that the prevailing norm in the mathematical social sciences is to build models that abstract from the

[1] Thanks are due Michael Cohen for stimulating my interest in this subject, and to participants in the Program for the Study of Complex Systems at the University of Michigan, for many useful comments.

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 33-48, 2000.

details of cognition.[2] Stated differently, the focus of economists and other quantitative social scientists on behaviorally simple models is a symptom of the lack today of anything like a universal model of cognition. A second reason for differential interest in models composed of moderate or large numbers of simple agents is that such systems are quite capable of complex aggregate behavior, involving, for example, the spontaneous emergence of behavioral norms (e.g., [19]) or the self-organization of multi-agent coalitions (e.g., [3]). Understanding the origin of these complex patterns is often a significant challenge, and would be even more difficult if individual agents were complex in their own right—if individual decisions were also emergent.

Given the relative simplicity of individual agents in such systems, it is almost certainly true that model specifications beyond the individual level play a somewhat more important role in such models than in MAS involving few agents. In particular, the interconnections between agents—the interaction topology—and the relative amount of individual activity in the agent population—the agent activation regime—surely must matter in wide varieties of models, especially to the extent that such models have empirical ends.

In this paper it is demonstrated that these factors—interaction topology and activation regime—can be crucially important in multi-agent systems, illustrated through a variety of empirically-oriented models of social phenomena. Specifically, when structures of interaction and activation are systematically altered, the aggregate statistics produced by such models can vary substantially. The next section addresses interaction networks in two multi-agent models. First, an existing multi-agent model of retirement dynamics, in which social networks are random graphs, is modified to have lattice-type networks in the space of age cohorts. This changes the overall behavior of the agent society as measured by the time required for establishment of a social norm in retirement age. Second, in a model of endogenous firm formation, where agents learn of alternative employment opportunities through social networks, a key statistic of the model is shown to depend crucially on the structure of these networks. Then, in section three we study the effect of two asynchronous agent activation regimes: *uniform activation*, in which all agents are active *exactly* once each period, and *random activation*, in which agents are active once a period *on average*. The effect of changing regimes is described in multi-agent models of trade, cultural transmission, and firm formation. A final section draws conclusions.

2 Effect of Agent Interaction Topologies

Social networks play a critical role as the medium within which human beings are socially situated and through which interactions between individuals occur. Therefore, it is hardly surprising that positive models of human social processes should include such networks. What is perhaps surprising is the extent to which relatively small changes in network structure can lead to large changes in macro-social outcomes. Experiments with human subjects involving systematic changes to

---

[2] There are many reasons for this state of affairs. The fact that cognitive psychology has focussed little on economic behavior [15] is certainly one reason, remediable in practice. That most economists have little training in cognitive science represents a larger challenge.

social networks are difficult to perform, of course. Here we utilize multiagent systems in their role as social science laboratories, since such models allow us to methodically alter such networks and to then discover, by spinning the models forward in time, the overall effects of such alterations. In this section, two models—one on the dynamics of retirement norms and the other a multi-agent model of firm formation—have their social networks systematically altered. In the first case a model employing random graph networks is modified to a lattice configuration, and then to a so-called 'small world' graph.[3]

Recently, procedures for creating 'small world' graphs have been introduced in [17]. Analysis of such graphs demonstrate that they possess a form intermediate between regular and random graphs, and have many properties of real-world social networks. Specifically, start with a d-dimensional lattice having a fixed number of edges. Then, systematically break each edge with some probability, π, and reattach each broken edge to a random node. There results a graph having a well-defined sense of 'localness' as in a regular graph (i.e., agents who know each other typically have significant overlap in their social networks), as well as relatively short paths between any two agents in the graph, a characteristic feature of random graphs. These 'small world' graphs are thought to be relatively good models of certain social networks (cf. [11], [16]).

2.1 Coordination in Social Networks: A Retirement Model

In [5] a model of retirement behavior is described in which there is a population of agents of various ages, with each agent having to decide when to retire. The model abstracts from economic factors in attempting to explain a certain puzzle in the evolution of the modal age of retirement—namely the long lag between the last big change in benefits policy in the U.S. and the systematic change in the overall behavior of retirees. Specifically, in 1961 an 'early retirement' age of 62 was instituted by the Social Security Administration. Benefits received by age 62 retirees were reduced in comparison to those who retired at age 65, by an 'actuarially neutral' amount—typical people would be, it was thought, indifferent between retiring at any age between age 62 and 65. The puzzle is that in 1961 the modal retirement age was 65, and this remained so up through 1990. Only by 1995 had this shifted, rather abruptly, to age 62. Standard rational actor accounts of retirement decision-making have difficulty explaining the long lag between changes in benefits and responses in overall behavior.

Our model of the retirement process employs a heterogeneous population in which there are three kinds of agents:

1. 'rationals' retire at the earliest age permitted by law;
2. imitators play a coordination game in their social networks—within an agent's network, if the fraction of agents who are retired among those eligible for retirement exceeds a threshold, τ, the agent too retires, else it stays working;
3. randomly-behaving agents retire with some fixed probability, p, once they are eligible to do so.

---

3 An illustration of the importance of social network structure in several game theoretic models can be found in [13].

This simple model is capable of reproducing certain features of retirement data from the United States, particularly a relatively long lag time from the change in earliest retirement age until its establishment as a behavioral norm in the population.

During each period of model execution, each agent ages one year and gets to decide whether or not to retire (there is also some chance of dying). The behavior of the first and third agent types is straightforward, while that of the second type, the imitators, can be characterized using game theoretic notions.[4] Consider a population of A agents, in which the state of agent i is $x_i \in$ {working, retired}. Then the state of society is given by $x \in \{\text{working, retired}\}^A$. Agent i has a social network, consisting of a set, N_i, of other agents. Overall, the utility that agent i derives, U_i, from interacting with members of its network is given by

$$U_i(x) = \sum_{j \in N_i} u(x_i, x_j), \qquad \textbf{(1)}$$

where the $u(x_i, x_j)$ can be thought of as payoffs in a 2 x 2 game, as in table 1.

Table 1. Imitation as a coordination game

	working	retired
working	*w, w*	0, 0
retired	0, 0	*r, r*

An agent's imitation threshold, τ, can be stated in terms of these payoffs, given the graph weights in its social network. In the case of uniform weighting—i.e., the behavior of all neighbors considered equally—$\tau = w/(r+w)$; for $r = w$, $\tau = 1/2$, and agents for whom the majority of their social network are retired will also retire. Furthermore, if the payoffs are made heterogeneous in the population then the imitation thresholds become heterogeneous as well.

For the imitators in the model, their networks are random graphs in [5].[5] Note that although each agent has a fixed network, the overlapping generations character of the population, in which old agents die and new, younger agents are born, renders the overall networks in society as transient. Here we study the effect of first moving to regular graphs of a certain type and then to small world graphs. Since human social networks tend to be highly correlated with age—i.e., people's designations of their friends are often dominated by people within a few years of their own age—the regular graph we employ is localized in the cohort space. That is, each agent has friends who are near it in age.

The key statistical output of this model is a measure of how long it takes a social norm of age 65 retirement to establish itself from an initial condition of no retirement age norm. In [5] the way in which this measure depends on various parameters of random graphs—e.g., their size, their extent in the cohort dimension, their heterogeneity—is studied. Here we investigate how altering the network structure to lattice social networks in the cohort dimension modifies the time required for establishment of this social norm. When populations are homogeneous and interactions occur on regular graphs, it is known analytically that the transition time

4 Formal results for games of this type played on static graphs are developed in [20].

5 For analysis of the properties of random graphs see [6].

between equilibria can be sped up through local interactions [7]. But for heterogeneous populations on transient graphs little is known analytically about waiting times, the lifetimes of transients, and so on.

In order to facilitate comparisons, the size and extent of agent networks is kept fixed as we vary network type. These and other parameters are described in table 2. The agents are parametrically homogeneous here, and have the same decision rule, but each agent has a unique social network and therefore the realized behavior in the population of agents is heterogeneous. These parameters differ somewhat from the 'base case' described in [5], involving larger social networks that have less extent in the cohort dimension ('network age extent' of 2 means that an agent's network includes agents who are within 2 years of its own age).

Table 2. Configuration of the retirement model

Parameter	Value
Agents/cohort	100
Imitation threshold, τ	50%
Social network size	24
Network age extent	2
Random agents	5%
p	0.50

This set of parameters yields a model in which the typical time required to establish an age 65 retirement norm is somewhat greater than in the base case in [5], due to the larger social networks. The rightmost line of figure 1 below gives these times as a function of the fraction of rational agents in the population, holding the number of randomly behaving agents constant at the level of table 2. Note that for approximately 15% rational agents, the average transition time is around 100 years. The bars above and below the average values in the figure represent ±1 standard deviation; these are asymmetrical because the ordinate is in logarithmic coordinates.

Running this model demonstrates that the state of being 'retired' percolates upward from below—from older cohorts—as it were.[6] The model behaves as if retirement were *diffusing through* or *infecting* the population, from older to younger individuals. In effect, the agents are an *excitable medium* through which retirement behavior can spread, with agent interaction serving to speed the adoption of retirement while at the same time the aging of the population (and the inherently transient nature of the overall social network) acts in the opposite direction, to limit adoption. Overall there is continual ebb and flow of the retirement state through social networks, until eventually it takes hold among essentially all agents capable of being in such a state. This uniform adoption of retirement behavior is best understood as the emergence of a social norm of uniform retirement age. While such norms, once established, can be destabilized by chance events, in general they are quite robust.

---

[6] A typical realization of this model can be found in QuickTime® format at www.brook.edu/es/dynamics/papers/interaction.

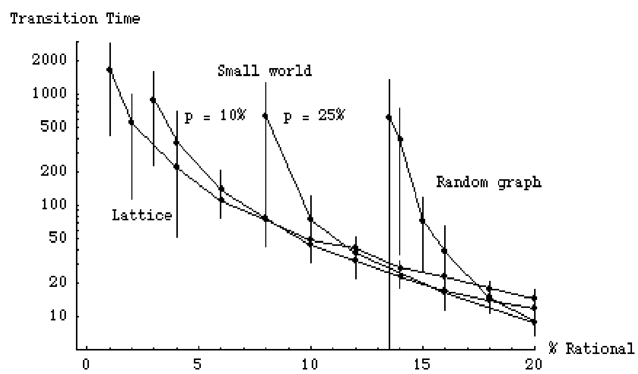


Fig. 1. Time until establishment of age 65 retirement norm, as a function of social network configuration

These results can be compared with those that obtain for lattice-type regular social networks. These are also depicted in figure 1, as the leftmost line. For a given fraction of the population acting rationally, lattice networks clearly require significantly less time for social norms to arise than random graph networks. The effect is dramatic at small levels of rationality, where random graphs would essentially take forever to lock in to a social norm, but lattice networks may require only a few decades. Equivalently, for a specified transition time a much smaller fraction of the population needs to be rational when social networks are lattice-like.

Typical model realizations[7] reveal that this model has a different dynamic for the establishment of retirement norms. In contrast to retirement percolating from older to younger cohorts, with lattice social networks retirement behavior starts among age 65 agents, first in a small group, and grows outwardly through that cohort as well as down through the population as it ages. Visually it is clear that this is a very different mechanism than with random graph networks.

As has been argued briefly above, neither random nor regular graphs well-represent empirically-significant social networks. In random graphs there is no sense of location, while in lattices the 'social distance' between two agents can be very large. Rather, real-world social networks seem to have the 'small world' property, i.e., networks are localized yet the path length between any two individuals is not large. From [17] we know there is a well-defined sense in which we can move from lattices to random graphs by randomly selecting links to be broken and randomly reattached. Such a process yields small world graphs. Figure 1 also reports results for the dynamics of retirement in such small world social networks, with two different values of the probability of breaking and randomly reattaching a link, π. For $\pi = 10\%$, the social networks retain their lattice look, but now there exist much shorter paths between any pair of agents. The establishment of retirement norms looks much like the lattice social network model in this case, although the transition times are somewhat longer. In the case of $\pi = 25\%$, the paths between arbitrary pairs of agents

7 Also available at www.brook.edu/es/dynamics/papers/interaction.

are even shorter. Here, the fraction of rational agents necessary to achieve a specified transit time to a retirement age norm is about halfway between the pure lattice and random graph social network cases. Clearly, these small world graphs behave as an intermediate form between regular graphs and random ones.

2.2 The Emergence of Firms

In [3] a model is developed in which heterogeneous, self-interested agents form groups in a team production environment with increasing returns. Each agent has preferences for income, gained from working, and for leisure (all time not spent working). Nash equilibrium effort levels in a group are Pareto-dominated by higher effort levels, although these are not individually-rational. The main analytical result for this strategic situation is that there is a size beyond which any group is unstable. That is, there is a maximum stable group size for any distribution of preferences. As groups exceed their maximum size, agents are permitted to join other groups or to start up new groups if it is welfare-improving to do so. It turns out that meta-stable groups—temporary coalitions—of agents can survive in this model out of equilibrium for significant lengths of time. We have studied such transient groups via a multi-agent system.

Interestingly, many of the statistical features of these groups closely resemble what is known about the population of firms. In particular, the size distribution of such groups is highly skewed, approximating a power law (Pareto distribution) in the upper tail, a well-known property of firms in industrial countries. Second, the growth rate distributions in the model are closely related to those found empirically for U.S. firms—essentially, non-normal distributions with heavy tails. Third, the way in which growth rate variance depends on firm size in the model can be made almost identical to the data (more on this later). Fourth, it is something of an empirical puzzle that wages tend to increase with firm size. This phenomenon can be found in the model, but only for particular interaction topologies. To understand how this is so, it is necessary to consider the details of agent decision-making.

When an agent is activated in this model, it assesses how its utility could be increased by altering its effort level. Perhaps several new agents have joined the group since it last re-evaluated its effort, or maybe other agents have systematically altered their effort contributions to production. This assessment by the agent could involve utility maximization, taking as given other agents' behaviors or taking into account the reactions of others to its own change in effort level. Alternatively, it could be simply utility-improving through a process of groping for better effort levels. However potential utility increases are determined, the agent stores these new efforts and utilities for comparison with other options, including joining other firms as well as starting up a new firm on its own. For each of these options the agent determines effort levels that improve its utility. But which extant firms do agents consider? There are many ways to do this. First, an agent could simply pick a firm at random in the population of firms. Alternatively, it could pick an agent at random from the overall population and consider joining its firm. Similarly, it might carry around with it a social network of 'friends' and each period consider joining the firms of its friends. Or, perhaps most realistically, firms that can profit most from hiring could post 'ads' in a virtual newspaper in order to attract potentially interested agents.

Clearly, these are quite varied ways of selecting prospective employers, and there is no obvious reason why they should yield the same results, especially given the highly skewed distribution of firm sizes. That is, due to the size distribution skewness, sampling a random firm is very different from selecting an agent at random. The former process produces a small firm with high probability—median firm size is under 10, the mode is 1 or 2—while the latter more frequently samples larger firms. Interestingly, it turns out that most of the empirical features of this model are robust, qualitatively, to such variations, i.e., size distributions remain skewed, growth rate distributions have fat tails, and growth rate variance scales with size. However, one empirical feature is sensitive to the structure of interactions, the wage-size effect. In figure 2 the dependence of wages on size is shown for two kinds of networks, one in which random agents are selected (lower line), and one based on choosing new firms (upper line). It is clear from this figure that there is little wage-size effect in the former case (i.e., random agents), while wages increase with size in the latter case. Empirically, wages are approximately proportional to size$^{0.10}$, while the upper line in figure 2 describes an almost identical relationship.

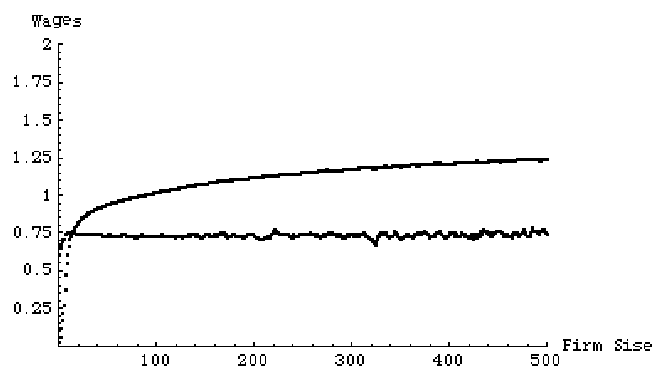


Fig. 2. Effect of firm size on wages, search networks based on new firms (upper) and random agents (lower line)

This clear difference in wage-size effects that results from changing the networks in which search is conducted is due to the different kinds of prospective employers yielded by the sampling process, as alluded to above. Due to increasing returns, growing firms in this model have high output per employee (productivity) and create high utility for their workers. When job-seeking agents can 'see' such firms, as with firm-based search networks, then output per worker grows with size. Alternatively, if such firms are infrequently sampled, as with random agent networks, then the 'arbitrage' in marginal utility that takes place in this model as agents migrate between firms yields constant output per agent across firms.

3 Effect of Agent Activation Regime

By agent activation we refer to the order and frequency of agent action. Because real social processes are rarely synchronous, it is rarely appropriate to model them by permitting each agent to update its states exactly once each period, using last period's state information.[8] Two types of asynchronous activation regimes are commonly employed in the large population-simple agent MAS described in this paper. In *uniform activation* a period is defined as the time during which all agents are active once and only once. In *random activation* each agent has an equal *a priori* probability of being active each period, but stochastic variations lead some agents to be more active than others. We study these in turn below.[9] In general, we can expect these alternative execution regimes to yield different individual agent histories, and possibly different macro-social outcomes as will be illustrated below. Indeed, even for a given type of activation it is only under very specialized conditions that distinct agent updating histories will yield invariant histories of agent states [9].

3.1 Uniform Activation

In uniform activation, each agent is activated once per period. A possible problem with this execution regime is that if agent *i* always acts immediately before agent *j* over the course of many periods then there exists the possibility that correlation between these agents will develop that is unrelated to the agent behavioral rules, but is rather a direct consequence of the activation structure. We call this *spurious* agent-agent correlation and suggest that in most cases it should be considered a programming artifact—an unintended consequence or side effect—and avoided. Happily, by randomizing the *order* of agent activation from period to period it is usually possible to remove all artifacts of this type and thus avoid any spurious correlation. Stated differently, with uniform activation it is crucial to periodically randomize the order of agent updating.

But how much randomization is appropriate? That is, given the order in which agents were serially activated last period, how many agents need to be repositioned in this sequence so that in the next period most of the agents either precede or follow a different agent—i.e., most agents have one or more new neighbors in the execution sequence? We have tried to answer this question analytically, without success.[10] To get some feeling for the difficulty of this question, consider the simple case of a single pair of randomly selected agents who are swapped in the agent activation list. How many agents will have 1 new neighbor? How many will have 2 new neighbors? There are three cases to consider:

1. The agents selected to be swapped are neither neighbors nor have any neighbors in common, so rearranging them will give each of their 4 previous

---

8 An important exception to this statement are multi-agent models of traffic. In [10] a critique of the use of synchronous updating in certain game theoretic models is rendered, demonstrating that results from such models are typically not robust to asynchronous updating.

9 The case in which agents have differential incentive to be active is treated in [12].

10 This problem is very similar to the card shuffling problem—how much shuffling is sufficient to satisfactorily "mix" a deck?

neighbors 1 new neighbor. Therefore, the fraction of the population that has exactly 1 new neighbor is 4/A, and each of the agents who moved has 2 new neighbors.

2. The agents to be swapped are immediate neighbors, so the process of rearranging them yields 4 agents who have exactly 1 new neighbor, and no agents with 2 new neighbors.
3. The agents to be moved have 1 neighbor in common, so the rearrangement process also produces 4 agents with exactly 1 new neighbor, and no agents with 2 new neighbors, although the common neighbor has the order of its neighbors reversed.

Now, to figure out the probabilities of having exactly 1 or 2 new neighbors it is necessary to determine the relative probabilities of these 3 cases. Then, the case of rearranging 2 agent pairs (4 agents total) involves 3 x 3 cases: the first agent pair generates the three cases above, and for each one of these the next agent pair creates 3 more cases. For an arbitrary number of rearrangements, this quickly becomes a very messy analytical problem and so we have resorted to a computational analysis, described in the following example.

Example*: Agent list randomization*

Consider a population of size *A* maintained in a linear data structure with the last element connected to the first, so that each agent has two immediate neighbors. The agents are activated according to their position in this list. We are interested in how much agent rearrangement is necessary to produce a well-shuffled list. Pick *L* agents at random and reposition them in the data structure.[11] What fraction of the agents in the list have at least 1 new neighbor? What fraction have 2 new neighbors?

This process has been studied for *L* varying to 150% of *A*—that is, from 2 agents up through 1.5 *A*—for various population sizes, *A* = 100, 500 and 1000. For each (*L*, *A*) pair, 100 realizations were made and statistics computed concerning the number of new neighbors. The average results did not differ appreciably across population sizes, and shown in figure 3 is the *A* = 1000 case.

Notice that for relatively few agent rearrangements, a relatively large fraction of the agents get at least one new neighbor—e.g., 25% rearrangement corresponds to nearly 50% of the agents having one new neighbor. However, it is also the case that in order to guarantee that 90% of the agents have 2 new neighbors, a very large number of rearrangements have to be performed—a number larger than the entire population! The lesson here is that random agent selection and repositioning is a relatively good way to accomplish modest rearrangement of the agent population list, but is too expensive to serve as a general method, especially when it is desired to completely rearrange a list. (If complete list rearrangement is desired then there are a variety of ways to accomplish this at less cost than random repositioning.) Summarizing, in those cases where having all agents active in a single period is behaviorally reasonable, i.e., uniform activation, figure 3 can be consulted to guide decisions about the amount of agent rearrangement to utilize each period.

[11] There are many efficient ways to do this, depending on the data structure involved, e.g., for fixed size structures such as arrays, swapping the positions of two agents consecutively drawn .

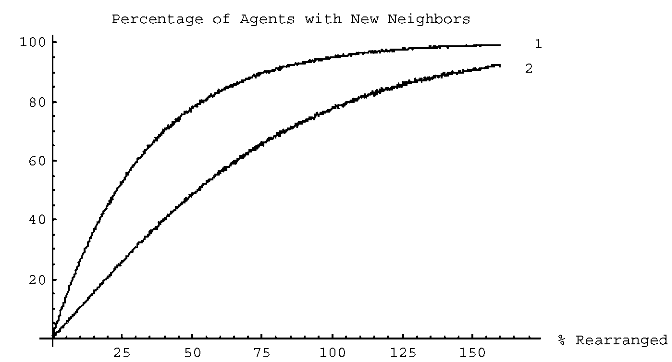


Fig. 3. Probabilities of having at least one (upper line) and exactly two new neighbors (lower line) as a function of the percentage of agents who are rearranged

3.2 Random Activation

Now consider the case of k distinct agents selected from a population of size A in order to interact. For $k = 1$ the agents are solitary actors, while for $k = 2$ these are bilateral interactions. When the probability that an agent is active is uniform throughout the population, the distribution of agent activation is binomial, and we call this *random activation*. In a population of size A, the probability that an agent is active at any particular instant is k/A. Over T instants the probability that an agent interacts each time is $(k/A)^T$, while the probability it does not interact at all is just $(1-k/A)^T$. Overall the probability that an agent is active i times over T instants is simply

$$\binom{T}{i}\left(\frac{k}{A}\right)^{i}\left(\frac{A-k}{A}\right)^{T-i} . \qquad (\mathbf{2})$$

The mean number of activations is kT/A, the variance is $kT(A\text{-}k)/A^2$, the coefficient of variation, $\sigma^2/\mu = (A\text{-}k)/A$, is independent of T, and the skewness coefficient is $(A\text{-}2k)/•(kT(A\text{-}k))$; note that this last quantity is positive for $A > 2k$. The time a particular agent must wait to be first activated—the waiting time—is a random variable, W, having a geometric distribution; its pmf is

$$\Pr(W = T) = \frac{k}{A}\left(1 - \frac{k}{A}\right)^{T-1} . \qquad (\mathbf{3})$$

The expected value of W is A/k, with variance $A(A\text{-}k)/k^2$.

For $T \approx A \gg k$, the average number of activations is approximately k, the variance is also about k, and the skewness lies around $1/•k$. Since the mean and variance are nearly equal in this case, many agents will fail to be active over time T. This can also be seen from the waiting time, the mean value of which is large with the variance approximately equal to the mean squared. A situation of this type can be viewed as problematical in multi-agent systems, where a reasonable presumption is that models are run long enough for all agents to be active.

From the waiting time distribution we can explicitly compute the probability that W exceeds T for a particular agent. Calculations of this type are shown in figure 4.

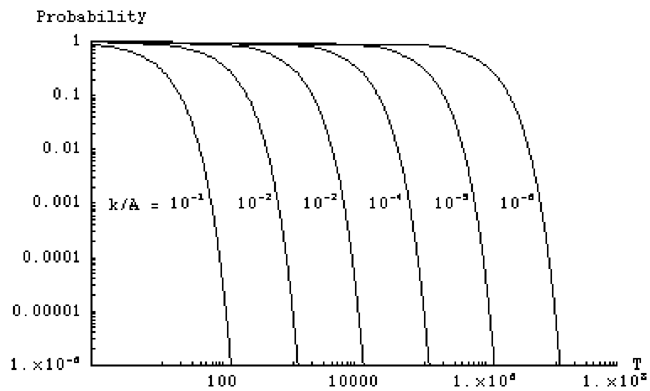


Fig. 4. Probability that a particular agent in a population of size A has been inactive over T activations when k agents are activated at once

As T increases beyond A, that is, T » A, the mean number of interactions rises together with the variance in proportion to T, while the skewness coefficient approaches 0; for A » k skew vanishes like 1/•(2T).

***Example**: Distribution of the number of interactions per agent in bilateral exchange processes*

Consider an economy in which agents are randomly paired to engage in bilateral exchange, a single pair of agents trades each instant (k = 2), and the probability that an agent is part of a trading pair is uniform across the population. For 100 agents (A = 100) over 1000 activations (T = 1000) the probable number of activations per agent is shown in figure 5.

The mean number of activations in figure 5 is exactly 20, although the modal event of 20 activations is only slightly more probable (p = 0.0897) than 19 (p = 0.0896). The standard deviation of this distribution is 4.43, and it is skewed slightly, with a 'fatter' right tail, as evidenced by a skewness coefficient of 0.217.

Clearly, uniform and random activation represent very different models of agent activity. In a certain sense, uniform activation is the zero variance limit of random activation—the number of activations per agent has no variance in the case of uniform activation. In the next subsection the differential effects of these two activation regimes are compared. But before moving on, a third activation regime will be briefly described.

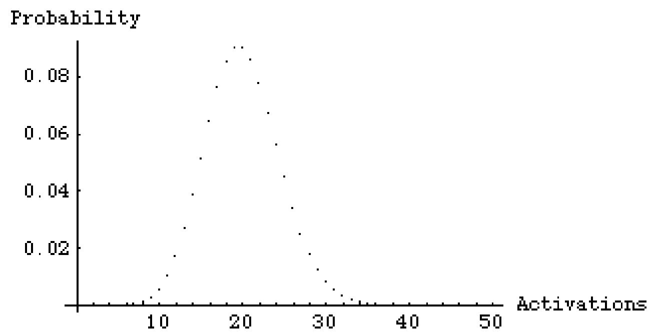


Fig. 5. Probability mass function for the number of agent activations in a population of 100 agents over 1000 activations when agents are paired sequentially with uniform probability

Imagine that each agent has its own 'Poisson clock' that wakes it up periodically in order to be active, such that the probability of its being active over a single time period is *k/A*. Then, over time *T* each agent is active *kT/A* times on average, just as in random activation, with the variance equal to the mean. Here, however, during any period of time the total number of agent activations is a random variable, and the random activation model described above is a kind of zero variance limit of random activation of the Poisson variety. There is a clear connection between these two types of random activation, since the Poisson distribution closely approximates the binomial for large values of *A* and *T*.

3.3 Comparison of Activation Regimes

The so-called Sugarscape model [8] is a multi-agent system designed for the study of demography, economics, disease propagation, and cultural transmission, on spatial landscapes. It utilizes uniform activation with execution order being randomized each period. A related but more elaborate model for the transmission of culture that uses random activation has been studied by Axelrod [2]. A 'docking experiment' was undertaken in which the culture component of the Sugarscape MAS was extended to incorporate Axelrod's model, in hopes of being able to reproduce the latter's somewhat counter-intuitive results. As described in [4], initial attempts to 'dock' the two distinct multi-agent systems yielded qualitatively similar results despite different activation regimes. However, statistical tests rejected the hypothesis that the data from the two models were the same. Only by altering the Sugarscape execution model to random activation was it possible to quantitatively 'dock' the two models. So this is an example of agent activation regime having a measurable statistical effect, although not altering the qualitative character of the output.

Returning now to the model of firm formation proposed in [3], one of the most striking regularities in the empirical data is that the variance in log(growth rates) decreases with firm size. Large firms simply have less variation in their growth rates than do small firms. Figure 6 below reproduces a figure in [3] that describes the close quantitative agreement between the model and the data on U.S. firms—essentially, the standard deviation in the logarithm of growth rates is a power law in firm size, with exponent approximately equal to -1/6. (The data scatter at large firm sizes is due to small sample sizes.)

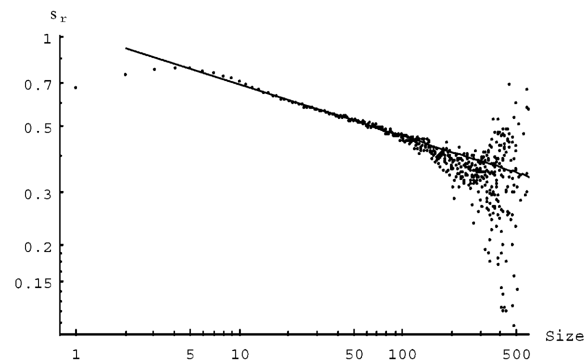


Fig. 6. Effect of firm size on the standard deviation in the logarithm of growth rate with random activation

The basic version of this model uses random activation. In the course of investigating the effect of alternative activation regimes on the model it was discovered that uniform activation (with 50% randomization each period) generates the reverse dependence of growth rate on size—variance *increases* with size—as shown in figure 7.

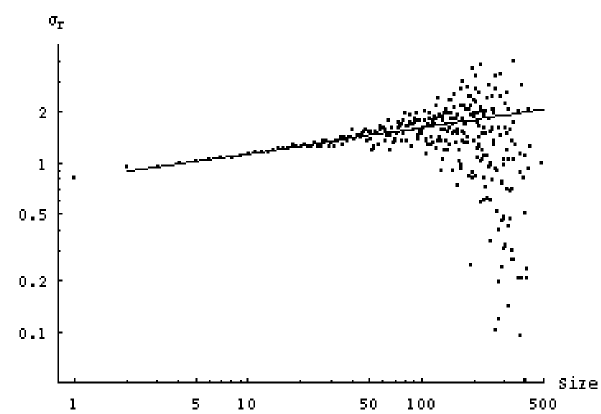


Fig. 7. Effect of firm size on variance in the logarithm of growth rate with uniform activation, 50% randomization each period

Here the slope is approximately +1/6, just the inverse of random activation. Such data could hardly be in greater disagreement with the facts—if true, General Motors would have more annual growth rate variation than some Silicon Valley start-up! The reason why this happens under uniform activation is that when large firms begin to decline—and it is a theorem in the model that all firms eventually decline—the fact that all agents observe this serves to hasten the process. Once a few agents realize there are better opportunities elsewhere, all other agents in the firm quickly react to this under uniform activation. Such 'unraveling' processes are more diffuse with random activation—those agents inactive during a period of decline in their firm look up from their desks one day to find things in bad shape, and only then do they begin searching for other positions. The random activation model implies more heterogeneous rates of sampling of external job opportunities among the agents within a firm.

4 Conclusions

The effects of agent interaction topology and agent activation regime have been investigated in several multi-agent systems. In moving between regular graphs (lattices) and random graphs, through small world-type graphs, the overall behavior of a model of the timing of retirement changed significantly. Then, altering the qualitative character of social networks in an empirically-accurate model of firm formation caused the wage-size effect in the model to disappear.

Two distinct agent activation schemes were compared and contrasted, uniform and random activation. These produced qualitatively similar but statistically different output in models of cultural transmission. In the firm formation model, random activation yields empirically-significant results. Moving to uniform activation generates qualitatively different and unrealistic output.

Clearly, agent interaction and activation structures can play important roles in MAS. Careful consideration of these 'architectural details,' including systematically studying how model output changes as such structures are varied, must be a key to robust analysis of multi-agent systems.

References

1. Aaron, H., ed. 1999. *Behavioral Dimensions of Retirement Economics*. Russell Sage Foundation/ Brookings Institution Press: New York/ Washington, D.C.
2. Axelrod, R. 1997. "The Dissemination of Culture: A Model with Local Convergence and Global Polarization." *Journal of Conflict Resolution*, **41**: 203-226.
3. Axtell, R.L. 1999. "The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distributions." Working paper 03-019-99. Santa Fe Institute: Santa Fe, N.M. Available at www.brook.edu/es/dynamics/papers.
4. Axtell, R.L., R. Axelrod, J.M. Epstein and M.D. Cohen. 1996. "Aligning Simulation Models: A Case Study and Results." *Computational and Mathematical Organization Theory*, **1** (2): 123-41.
5. Axtell, R.L. and J.M. Epstein. 1999. "Coordination in Transient Social Networks: An Agent-Based Computational Model of the Timing of Retirement." In Aaron, ed. [1]. Available in working paper form at www.brook.edu/es/dynamics/papers.
6. Bollobás, B. 1979. *Graph Theory*. Springer-Verlag: N.Y.
7. Ellison, G. 1993. "Learning, Local Interaction, and Coordination." *Econometrica*, **61**: 1047-71.
8. Epstein, J.M. and R. Axtell. 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press/Brookings Institution Press: Cambridge, Mass./Washington, D.C.
9. Gács, P. 1997. "Deterministic Computations whose History is Independent of the Order of Asynchronous Updating." Working paper. Boston Univ.: Boston, Mass.

10. Huberman, B.A. and N.S. Glance. 1993. “Evolutionary Games and Computer Simulations.” *Proceedings of the National Academy of Sciences USA, 90*: 7716-18.
11. Kochen, M., ed. 1989. *The Small World***.** Ablex Publishing Corporation: Norwood, N.J.
12. Page, S.E. 1997. "On Incentives and Updating in Agent Based Models." *Computational Economics*, **10** (1): 67-87.
13. ———. 1999. "Network Structure Matters." Working paper. Department of Political Science. University of Michigan.
14. Rubinstein, A. 1998. *Modeling Bounded Rationality*. MIT Press: Cambridge, Mass.
15. Simon, H.A. 1998. In Rubinstein [14].
16. Watts, D. 1999. *Small Worlds: The Dynamics of Networks Between Order and Randomness.* Princeton University Press: Princeton, N.J.
17. Watts, D.J. and S.H. Strogatz. 1998. “Collective Dynamics of Small-World Networks.” *Nature,* **393**: 440-442.
18. Weiß, G., ed. 1998. *Multi-Agent Systems*. MIT Press: Cambridge, Mass.
19. Young, H.P. 1998. *Individual Strategy and Social Structure*. Princeton University Press: Princeton, N.J.
20. ———. 1999. Diffusion in Social Networks. Center on Social and Economic Dynamics working paper. Brookings Institution: Washington, D.C. Available at www.brook.edu/es/dynamics/papers.

Simulating Emergence and Downward Causation in Small Groups

R. Keith Sawyer

Washington University in St. Louis, Program in Social Thought and Analysis,
Campus Box 1183, Saint Louis, MO 63130
ksawyer@artsci.wustl.edu

Abstract. Emergence has been a central issue not only in computational models of social systems, but also throughout 20th century sociological theory. In this paper, I first define the key concepts of emergence, downward causation, and bi-directional causation, and I identify these processes in an example of improvised dialog. I then draw on emergentist trends in sociological theory to propose a computational model of collaborative emergence among small groups of improvising agents.

1 Introduction

Many MAS simulations are *emergent systems*—complex systems that display behavior that cannot be predicted from a full and complete description of the component units of the system [26, 31]. Emergence has become an influential concept not only in MAS, but also in artificial life [30], contemporary cognitive science [10], complexity theory [4], and robotics [1].

Emergence is particularly important in those MAS that are *social simulations* [17,22]. Although few sociologists have used this terminology, emergence is also foundational in sociological theory, as a way of theorizing the relationship between the individual and the social [49]. For example, emergence is at the heart of the *micro-macro* problem: How are individual actions related to social structures [2, 28]?

Although MAS simulations are commonly said to manifest emergence, only a few researchers have attempted to define the properties of emergent systems, or to define different classes of emergence [3, 6, 13, 45]. These researchers suggest that at least two variables contribute to emergence: the number of units in the simulation, and the complexity of the communication among the units. For example, Darley [13] and Sawyer [45] proposed that emergence is a function of both the number of units and the complexity of the rules of communication.

Communication complexity varies dramatically in reactive agent and cognitive agent systems. Reactive agent simulations use quite simple communication languages. They display emergent properties not because of their communication complexity, but because of the large number of units. For example, Alife-derived social systems models use rules such as those proposed by Epstein and Axtell [17, p. 73]:

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 49-67, 2000.

Cultural transmission rule (tag-flipping):

- For each neighbor (4 orthogonally contiguous agents), a tag (one bit in an 8-bit mask) is randomly selected;
- If the neighbor's bit setting agrees with the agent's at that position, take no action; if they disagree, flip the neighbor's tag to agree with the agent.

Communication in cognitive agent MAS is much more complex, due to the sophistication of *agent communication languages (ACLs)*. Although cognitive agent MAS have many fewer units than reactive agent MAS, their increased communication complexity might contribute to a different form of emergence. However, as Castelfranchi noted [7], emergence has not yet been studied in cognitive agent MAS.

Sociological theories of emergence have begun to incorporate sophisticated theories of symbolic interaction [12, 39]. These microsociological studies suggest that simulations of social emergence must focus on cognitive agents with complex ACLs, because they have suggested that network connections, interaction patterns, and complexity of communication must be incorporated into macrosociological theory.

In addition, empirical studies have demonstrated that natural human communication is much more complex than any existing ACL. To date, ACL have been largely based on speech act theory and BDI agent models [38]. However, speech act theory has been broadly discredited by these empirical studies of conversational interaction [47]. Sociological studies of interaction and emergence suggest several elaborations of ACLs that can result in more sophisticated social simulations.

This paper represents an early stage in a long-term project: to generate simulations of collaborating groups, and to apply insights from these microsociological simulations to *macrosociological* theory—the theory of institutions and social structures. I refer to emergence in small groups as *collaborative emergence*, to emphasize several contrasts with patterns of emergence found in reactive agent systems [45]. Social group behavior is collaboratively emergent in those cases where there is not a structured plan guiding the group, and where there is no leader who directs the group. Examples of collaborative emergence include everyday conversation, small-group collaborations, brainstorming sessions, and discussion seminars. All of these phenomena are *improvisational*, because there is no director and no guiding script. Consequently, I found it necessary to precede my MAS modeling efforts by analyzing creative improvisational performances, including jazz, improvisational theater, and children's fantasy play [40, 43, 44]. These studies have identified many of the interactional processes that make the difference between a coherent emergent encounter and incoherent, unconnected discourse.

Improvisation is of particular interest, because it is in improvisational interaction that emergent processes are most likely to be prominent. Among situated robotics researchers, intelligent agent behavior is thought to emerge from the improvisational and contingent interaction of the agent with its environment [1]. For robotic agents acting in real-time, improvisation is thought to be necessary because many situational tasks are undecidable in advance [8].

1.1 Downward Causation

In addition to their relatively simple communication languages, there is a second limitation of most contemporary MAS that keeps them from adequately accounting for social emergence: their lack of downward causation.

A MAS simulation, although said to be emergent, can be fully explained in reductionist fashion. Drawing an example from Alife, an emergent pattern like a glider is fully explained by the rules and the initial state of the system. The glider doesn't exist; it only seems to be a coherent, moving object because of the way our perceptual system is designed. Such emergent patterns are epiphenomenal—they do not have causal force that is autonomous from the composing entities—and can be reduced to lower-level explanation [18, 21]. In MAS and Alife social simulations, the emergent pattern is fully explained by the microsimulation; that is, reduced to an explanation in terms of agents and their interactions. Such reductionist assumptions imply that higher-level emergent patterns do not have any causal force. Thus, emergence is simply a shorthand for describing higher-level patterns that are, ultimately, reducible to descriptions of agents and their interactions.

But in natural social systems, emergent patterns seem to influence both individuals and emergent processes. The status of this social *downward causation* has been a major theme in 20th-century sociology [11, 15, 37]. In small groups, for example, the emergent is the interactional frame of the conversation. The interactional frame is generated non-additively from the combined actions of participants; and in each turn of the conversation, the frame exerts constraints on the subsequent action [42, 45].

MAS have rarely been based on sociological theory; rather, they tend to draw on economics models that are based in either rational choice theory or game theory. In sociological theory, the assumption that social phenomena can be explained by reduction to individuals is called *psychological reductionism*, *methodological individualism*, or *reductionist atomism*. Contrasting schools of sociological theory—including sociological realism, sociological holism, and structural sociology—have attempted to theorize the social level of analysis in such a way that it can be argued to be, in principle, not reducible to the individual level of analysis. In this article, I draw on these latter sociological theories to elaborate the notion of emergence, in a way that is compatible with ontological individualism—groups are nothing but their constituent agents—and yet rejects methodological individualism, the claim that the only way to analyze and predict social phenomena is through reduction to individual-level theory and explanation. Such a conclusion would have significant implications for the design of social simulations, because to date they have largely been based on methodological individualist assumptions.

Downward causation has also been a major concern in late-20th century philosophy of mind; theorists have explored the possibility of downward causation in neural systems, as a way of theorizing how intentional mental states have causal effects. Many philosophers of mind argue that there can be mental causation, even within a materialist ontology that accepts that only physical matter exists. These philosophers argue that although the mental is grounded in the biological brain, the mental may nonetheless be real in the sense that it may have autonomous causal powers [14, 19]. Philosophers of biology [5, 51] likewise argue that there are processes of downward causation in biological systems, such that consciousness has a determinative influence on the underlying neurophysiological substrate. And in their studies of far-from-equilibrium systems, Nicolis and Prigogine [35] described a range of natural physical systems, maintained by large influxes of energy, in which higher-level structures exert downward causation on the component particles. By analogy, the social may have autonomous causal powers, even though it is grounded in collections of individuals and their interactions [49].

Downward causation has not been represented in MAS social simulations [cf. 21, 29]. Some MAS researchers have simulated phenomena that emerge and then remain stable over generations, using concepts of self-organization and positive feedback [53]. There is similar recent work in Alife systems [52]. Yet in all these systems, the higher-level emergents do not exert causation on individual agents; the causation is exerted only through adaptation in local interactions, and no agent has a representation of the macro pattern.

Agent-based models are often concerned with the agent's interactions with an environment, and many of the agent's rules indicate how to respond to features of the environment. Some simulations have modeled the networks of social connections as a part of the environment [32]. In such systems, features of the local social environment are modeled by agents, but none of the emergent global properties of the simulation are internalized by the agents.

When higher-level emergents take on causal properties, they take on what seems to be an ontological status independent of the components. But where does the apparent causal force of the emergent come from, if not from the lower-level interactions? Even if nothing new emerges, computational modeling may nonetheless be required to proceed *as if* emergent macro phenomena participated in causal laws, if the equivalent individualist reduction can be demonstrated to be non-computable [34] or undecidable [8].

For purposes of computational modeling, it is not necessary to take a position on philosophical issues such as whether or not the emergent really exists, or whether a reduction to the agent level is in principle computable or not. As with any other modeling of a natural phenomenon, one should attempt to model the observed regularities, which are typically found at multiple levels of analysis. The emphasis on reductionist atomism in contemporary social systems modeling is not based on empirical observation of social life, nor on unanimously accepted theoretical foundations, but is an artifact of a historical situation: (1) the fact that appropriately formalizable theories of group behavior have, to date, only been available in atomistic theories such as rational choice theory and game theory; (2) the availability, beginning in the 1980s, of parallel computer architectures and object-oriented languages, which provide a powerful new tool for exploring reductionist atomism.

1.2 Bi-directional Causation in Social Systems

In Alife-based social simulations, the emergent is usually considered to be some stable endstate of the simulation. Researchers vary properties of the agents, and then examine the different (visual) higher-level patterns that result. The interactional processes which result in the generation of this endstate have not been a central focus of Alife research.

In small social groups, in contrast, the emergent frame is not a stable end state, but is rather a constantly changing creation. Thus, most social emergents are constantly changing. For example, a language like English is emergent and collaboratively created [25], but it does not have an independent physical existence. And although it largely seems stable within the lifespan of a single individual, in fact there is constant evolutionary change.

In collaborative emergence in small groups—for example, the emergence of an interactional frame from dialogue—it is particularly obvious that the emergent interactional frame is a process rather than a product. At each stage of the simulation, the emergent frame is slightly modified, and in turn, that frame influences the next stage of emergence. Thus emergence and downward causation act in a tight feedback loop, resulting in complex and potentially chaotic behavior. In such simulations, the incremental process of emergence, rather than the endstate, is the focus [cf. 24].

There is a trivial sense in which all agent simulations exert processual causation: each action of an agent affects the future actions of the entire system. However, this is not downward causation from an emergent, because the emergent is not modeled distinctly; the only effects on an agent derive from local interactions with other agents. A few agent simulations have recently moved towards representing such bi-directional causation, by allowing the agents to represent relationships among individuals [e.g., 36]. However, these representations are not themselves emergent, but are part of the initial conditions of the simulation.

2 An Example of Collaborative Emergence: Improvisational Theater Dialogue

Improvisational theater dialogues display the essential characteristics of collaborative emergence. Actions of individual agents give rise to the incremental emergence of a dramatic frame; once it begins to emerge, the frame then constrains the later actions of agents. Consequently, improvised dialogue is an example of bi-directional causation in multi-agent interaction.

The transcript in Example 1 is taken from a performance by a Chicago theater group. This is the first few seconds of dialogue from a five-minute improvised scene. The audience was asked to suggest a proverb, and the suggestion given was “Don’t look a gift horse in the mouth.”

Example 1. Lights up. Dave is at stage right, Ellen at stage left. Dave begins gesturing to his right, talking to himself:

1	Dave:	All the little glass figurines in my menagerie The store of my dreams
2	Ellen:	*Slowly walks toward Dave*
3	Dave:	*Turns and notices Ellen* Yes, can I help you?
4	Ellen:	*Ellen is looking down like a child, with her fingers in her mouth* Um, I’m looking for uh, uh, a present

5	Dave:	A gift?
6	Ellen:	Yeah.
7	Dave:	I have a little donkey?
		Dave mimes the action of handing Ellen a donkey from the shelf
8	Ellen:	Ah, that's=
		I was looking for something a little bit bigger…
9	Dave:	Oh. *Returns item to shelf*
10	Ellen:	It's for my Dad.

By Turn 10, elements of the frame are starting to emerge. We know that Dave is a storekeeper, and Ellen is a young girl. We know that Ellen is buying a present for her Dad, and because she is so young, probably needs help from the storekeeper. These dramatic elements have emerged from the collective interaction and creative contributions of both actors. No single actor determines the direction of the scene.

In many cases, an actor cannot know the meaning of her own turn until the other actors have responded. In turn 2, when Ellen walks toward Dave, her action has many potential meanings; for example, she could be a coworker, arriving late to work. The complete meaning of a turn is dependent on the flow of the subsequent dialogue (below, I note the implications for BDI models of agent interaction).

The subsequent dialogue is constrained by the dramatic frame that has begun to emerge; individual actions must be consistent with this emerging frame. Because Ellen is now a young child and a customer, she cannot suddenly begin to act as a co-worker. This constraint is a form of downward causation. As the dialogue continues beyond Turn 10, we learn that Ellen is buying her Dad a present because he has not been feeling well; in fact, he has been exhibiting psychotic behaviors. A third actor enters the scene, enacting the character of Ellen's Dad, and his condition is cured through some clever actions by the storekeeper.

This dramatic frame—the characters, motivations, relationships, and plot trajectory—collaboratively emerges. No one actor's turn determines its structure, yet with each actor's line, one possible path is chosen, and many other potential paths are closed off. The actors also experience downward causation of the emergent frame: Their individual actions must be dramatically consistent with the frame that is active at each moment.

3 Sociological Theories Relevant to MAS

Sociological theory has been an active subfield of sociology since the 19th century, and has perhaps never been more active than at present. Given such a high level of activity it is not surprising that there is a wide range of sociological theories in addition to the individualism associated with MAS. I will group these non-individualist theories into four categories, based on how they theorize emergence and downward causation in social systems.

3.1 Holism

Holism is closely related to sociological positions known as *sociological realism* and *sociological positivism*. Holism holds that sociology must study social structures in their own right, independent of facts about the individuals participating in the structures. Sociological method must treat social structures (groups, institutions, symbols, values, norms, statuses, dynamic and historical processes) as if they are natural facts, on the order of the facts of physics or biology. Canonically, this position is associated with Emile Durkheim [15], and holist assumptions underlie much of 20th century sociological research. Contemporary sociology has identified many laws in which social structures seem to have causal effects.

Although holists sometimes invoke emergence as a way of ontologically grounding their claims, they generally do not propose specific mechanisms for emergence processes, since they hold that social structures have explanatory primacy over individualist explanations, and a specification of emergence processes can often seem to be reductionist.

The most common version of downward causation proposed by holist theories is *internalization*: agents internalize a representation of the (externally existing) social structure. Holist theories are typically not interested in the nature of this process, except to the extent necessary to explain social phenomena at the structural level itself. For example, the issue of the *reproduction* of the social order—how the structure remains stable over time—is thought to require that individuals internalize, through socialization processes, certain more-or-less accurate representations of that structure. The widespread acceptance of this notion in American sociology is largely due to the influence of Talcott Parsons [37].

3.2 Subjectivism

Subjectivism holds that social structures do not exist: only individuals exist, since groups are made up of nothing other than individuals. The sociologist should focus on individual's subjective perceptions and interpretations of the social structure. Thus, subjectivism is implicitly reductionist. Such theories are often called *action theories* because they focus on individual's actions, and the meanings and intentions attached to those actions. Canonically, action theory originated with Weber at the turn of the century; the approach became more subjectivist after World War II [54], drawing on both phenomenology and sociological theory.

Subjectivist theories do not offer much of interest regarding either emergence nor downward causation, since they typically focus on a single individual. However, subjectivism is of historical interest because it has contributed to a range of contemporary *hybrid theories* which attempt to extend holist models by incorporating intentional states and subjective interpretations of participating actors.

3.3 Hybrid Theories

In the latter decades of the 20th century, a dominant trend in sociological theory has been to reconcile subjectivist and holist theoretical perspectives. Perhaps the pre-eminent theorist advocating such a reconciliation has been Anthony Giddens [20]. Many such theories take on an emergentist tinge, although none have made this explicit. For example, Giddens speaks of *situated practices*, emphasizing that they are different from isolated practice and thus implying emergence; but he does not empirically examine interaction, and in fact he explicitly rejects emergence (pp. 169-174). The most relevant aspect of this work is that it has developed a more sophisticated notion of downward causation than traditional holist theories: Rather than a straightforward, iconic internalization of the social structure, the downward causation is considered to be filtered through the individual's intentional states. This version of downward causation is compatible with MAS approaches to social reasoning, including BDI systems in which agents model the beliefs and desires of other agents.

Hybrid theories are attempts to address two issues that are central in current sociological theory: the relations between structure and individual agency, and the relations between micro- and macro-sociology.

3.4 Interactionism

Unlike the above theoretical traditions, *interactionism* holds that symbolic interaction among agents should be a primary focus of sociological research. Interactionist theorists have proposed that a focus on situated interaction can subsume micro and macro concerns (12, 28). This approach grew out of sociological sub-fields including symbolic interactionism, ethnomethodology, conversation analysis, sociocultural psychology, and critical discourse analysis. This shift in focus is of particular interest to MAS, since a focus on interaction seems to necessarily lead to a concern with emergence processes.

However, interactionist sociologies have not yet elaborated their theories in this direction. Typically, interaction is treated as an isolable level of analysis, and the relations with other levels are neglected—whether the individual's cognitive representations underlying the interactional behavior, or the higher-level emergent social structure that results from interaction.

Interactionist theories are of particular relevance to designers of ACL, since interactionist sociologists have developed quite sophisticated theories of natural language communication, moving far beyond the speech act/BDI models typically used in MAS.

3.5 Emergence in Sociology

Although none of the above sociological theories directly address the dialectic of emergence and downward causation, each of these theories provides elements that can contribute to a more sophisticated model of agent behavior in complex social systems.

- Holist theories have provided documentation of many forms of downward causation from social structure.
- Subjectivist theories have demonstrated the importance of considering agent's intentional states, when examining their action within social structures.
- Interactionist theories have documented the forms of sophisticated interactional behavior that mediate both emergence and downward causation, as agents engage in collective action within complex social systems.

The next step for sociological theory is to combine these elements with insights from philosophy and from MAS, to develop a complete theory of emergence and downward causation in social systems. Thus, MAS is poised to offer sociological theory exactly what it most needs at this point in its development. At the same time, this relationship could be mutually beneficial, since sociological theories offer many of the sophisticated components that are most lacking in social simulations.

4 Implications for Computational Modeling

Sociological research indicates that natural social systems are radically more complex than current social simulations. Of course, the unanswered question is whether the difference in complexity is quantitative or qualitative: Can current simulations ramp up to complex societies, without fundamentally changing their base assumptions? In this section, I briefly review relevant computational techniques, and for each, I identify limitations of the approach that may make it difficult to ramp up to such complex social systems. In section 5, I propose some computational extensions to MAS that could address some of these inadequacies.

4.1 Artificial Life

Alife models are lacking several of the features of natural social systems. They are characterized by:

- Agents that are not capable of maintaining internal representations of global structures
- A radically simplistic communication language
- No representation of the emergent social structure

Although Alife models are said to manifest emergence, these emergent higher-level patterns are epiphenomenal, and have no causal effects on the agents. This is quite different from the bi-directional causation observed in natural social systems.

4.2 Multiple Agent Systems

Most agent development tools are not ideal for small-group simulation, because they assume various structural features that tend to preclude emergence and downward causation. For example, most agent systems rely on a message-passing mechanism of point-to-point communication. The sender determines who should receive the message; and, once received, the recipient executes the message by choosing a corresponding method. However, in conversational groups, messages are multicast, and each recipient decides how to interpret the message, whether or not it is appropriate to respond, and to what extent to respond. In Example 1, it is exactly this contingency and intersubjectivity that enables collaborative emergence.

Like Alife, MAS have not yet attempted to explicitly represent emergent phenomena, nor have agents been developed that can internalize representations of emergent global social structures. Like Alife, then, MAS do not yet model the bi-directional causation that is found in natural social systems [cf. 24, p. 154].

4.3 Agent Communication Languages

Interactionist sociologies have many implications for developers of ACLs. Although current ACLs are perhaps sufficient for the unambiguous, goal-directed behaviors desired of engineering applications, they are not able to model natural human communication in social settings. Thus, such agent languages may not be capable of manifesting true emergence—in the sociological sense of generating an emergent, higher-level structure that can then influence the future behavior of agents.

Intersubjectivity has been recognized by linguistics researchers as the *common ground* problem: how do agents establish that they all share the same representation of what is going on [9]? However, ACL implementations do not address these issues; existing ACLs (e.g., KQML, FIPA) do not have any ambiguity that would require the agents to be capable of a grounding process.

All of the predominant ACL standards are based on speech act theory, and language specifications consist of a list of supported speech acts. Speech act communication typically goes hand-in-hand with agent models that use a Belief-Desire-Intention (BDI) framework [38]. "Beliefs" correspond to information the agent has about its environment; "Desires" represent goals from which the agent can choose; and "Intentions" represent Desires which have been chosen and committed to. Yet recall Example 1, where both environmental information and the meaning of an action are collaboratively negotiated by three actors. The existence of such natural communicative phenomena are problematic for intention-based models.

Because agent systems are based on unambiguous speech acts, intentionality and semantics are unproblematic. Thus there is no need to retrospectively negotiate the meaning of an action.

4.4 Blackboard Systems

The emergent social structure has some superficial similarities with *blackboard systems*, where all agents have both read and write access to a shared global database [16]. The blackboard can be considered to be an emergent product of the agents' collective actions, that then influences later agent actions. Yet blackboard systems have not modeled key aspects of social emergence: *intersubjectivity* and *internalization* of the emergent.

In small group interaction, the emergent is an interactional frame; outside of culturally-specified interactional encounters (such as Schank's canonical restaurant script) these frames are improvised and collaboratively created. Each participant internalizes a model, and due to the ambiguity present in many communicative situations, there may be slippage between these internalizations. This raises the issue of *intersubjectivity*: To what extent are individual agent's internalized models in alignment? How do agents handle potential slippage through their communication language?

Traditionally, intersubjectivity is defined as a state of overlapping, symmetrical mental representations; two or more people are said to "have intersubjectivity" when their mental representations of the situation are in agreement. This traditional view is implicitly reductionist, because intersubjectivity is reduced to individual subjectivities and their additive relations. In other words, intersubjectivity, and hence, all collective activity, is regarded as a simple sum of individual mental states [33]. Blackboard systems enforce this form of intersubjectivity, since there is only one representation of the emergent.

Yet there are many social interactions where participants do not share mental representations, such as disputes, arguments, and debates. In fact, even when there is no overt disagreement, it's unlikely that participants would have identical mental representations of what is going on. In the above improv theater transcript, there is a high degree of ambiguity at each dialogue turn. Although each actor may have a rather different interpretation of what is going on and where the scene might be going, they can nonetheless proceed to collectively create a coherent dramatic frame. The key question about intersubjectivity is not how agents come to share identical representations, but rather, how a coherent interaction can proceed even when they do not.

4.5 Interactive Environments

Several intelligent agent projects have drawn on metaphors of improvisation and theater, to guide their attempts to develop virtual reality and interactive environments; these include *Oz* [27], *Virtual theater* [24], and *Creatures* [23]. Perhaps the first of these was Bates's *Oz* project, an interactive media environment with rich interactive agents, with a focus on generating a story interactively. Unlike many agent systems, their concern with generating an interesting story line led them to consider emergence within a higher-level structure, a form of downward causation. They represented the structure using a *plot graph*; however, this is not emergent, but must first be created by an author. Yet, the simulation emerges within the constraints specified by the plot graph.

In Hayes-Roth's *Virtual Theater* project, the human user acts as a theater director, and the actors are animated agents capable of improvisational interaction. The human director thus is capable of downward causation, and the drama emerges from the interaction of the intelligent agents within these constraints. However, none of these emergent patterns, in turn, had causal effects, thus there was no bi-directional causation.

Such systems are of interest because they explore the improvisational actions of agents that are constrained by higher-level structures. Yet, these structures are not, themselves, emergent from agent interaction.

5 A Framework for Modeling Collaborative Emergence

Sociological theory and research suggests that natural social systems are radically more complex than current social simulations. Most agent development tools make several assumptions that tend to preclude emergence and downward causation: the emergence of an intersubjectively shared, contingent interactional frame is not modeled; downward causation of such a frame is not modeled; and agent communications are overly simplistic, based on speech act theory and BDI models. Yet recall Example 1, where both environmental information and the meaning of an action are collaboratively negotiated by three actors. The existence of such natural communicative phenomena are problematic for intention-based models.

A computational model of collaborative emergence requires the modeling of two features not found in MAS:

- Emergence occurs in parallel with downward causation, in a bi-directional dialectic. In improvised dialogue, for example, the emergent dramatic frame constrains future agent actions. Existing MAS simulations do not account for downward causation.
- In collaborative emergence, the communication system must be complex enough to handle intersubjectivity. In an improvised performance, for example, there is slippage among the mental models of the actors. In Turn 2 of Example 1, Ellen may have intended her approach to be that of a coworker arriving late, yet Dave addresses her as a customer. Consequently, for a coherent interaction to emerge, the actors have to be able to negotiate among their distinct representations. ACL must be extended so that the language can refer reflexively to itself, and the processes of communication themselves can be discussed.

5.1 Modeling the Emergent

As in blackboard systems, the emergent frame must be represented as a data structure external to all of the participating agents. In the case of improvisational theater, aspects of the emergent frame that must be collaboratively negotiated include the *characters* that each actor enacts (in Example 1, the storekeeper and the young girl); the *relationships* among these characters, including relative status and role; the *motives* of these characters, and how they attempt to achieve them; the *location* of the scene; the *shared activity* that the characters participate in; the *time period* and *genre* of the performance; and the *plot events and sequence* that are created during the improvisation.

There are several differences with blackboard models:

- All emergent collective structures must be internalized by each agent, resulting in an agent-internal version of the emergent. This internalization process is not deterministic and can result in each agent having a slightly different representation.
- The emergent structure's content will be a flexible conceptual representation, with each slot value having a range of specificity and ambiguity. By Turn 10 of Example 1, we know that Ellen is buying a present for her Dad, but we don't know the reason for the gift. The same is true of the agent-internal representations, although these are likely to be more specific and less ambiguous, due to cognitive limitations.

5.2 A Metapragmatic Communication Language

In improvised dialogue, every actor's turn both enacts a character within the frame, and at the same time, negotiates this shared understanding, by proposing an additional elaboration or transformation of that frame, and resolving intersubjectivity issues with other agents. This latter communicative function is *metapragmatic* [50], in that its indirect pragmatic effect is to further define the nature of the ongoing interaction itself. Improvised communications—like all human discourse—thus have effects on two levels: a speech act, or pragmatic level, and a metapragmatic level.

Collaborative emergence thus requires a complex symbolic communication system, one that supports both pragmatic and metapragmatic communication. Although such a system could be modeled using distinct media or channels for these two functions, human language manages to accomplish both functions simultaneously, using a single channel [50].

Developing a model of metapragmatic communication will require extending ACL by drawing on fields that empirically study naturally occurring discourse, such as sociolinguistics and conversation analysis.

5.3 The Incremental Process of Collaborative Emergence

In prior work [41, 42, 44] I have presented various components of a model of emergence in improvisation. Processes of emergence are characterized by *incrementalism*. At each step in the simulation, an actor can only modify the emergent by a small amount.

One agent acts at a time. An agent's act consists of the following steps:

1. Each agent internalizes the updated collective emergent.

- The internalized representations will necessarily be less complete than the collective one.
- The internalized representations will not be exact copies of the collective one; there will be variations due to ambiguity in the slot values (note that slot values will be more-or-less specific).
- The internalized representations may, at times, directly contradict one another, particularly among unskilled interactants. Children, for example, are not as talented at such interactions, and during dialogue, their internal representations are out of sync more often than adults [44].

2. Choosing the next agent to act.

Some actions will select the next agent, as when a question or comment is directed to a specific agent. Other actions allow any agent to act next. In this case, each agent has a random delay before deciding to act; the agent with the shortest delay makes a bid to act. In either case, the selected next agent constructs an improvisational action.

3. The improvisational action.

There are two pragmatic components of an improvisational action, a RESPONSE and an OFFER.

An OFFER is an incremental, creative addition to the emergent frame, such as a character, relationship, location, or joint activity. In Example 1, Dave's offer in Turn 1 is that he is a storekeeper, and the location of his action is his store; in Turn 4, Ellen offers a joint activity, shopping for a present.

The RESPONSE component relates to the status of the prior act. All OFFERs enter the collective frame provisionally; later agent actions can accept, partially accept, extend, reject, partially reject, or modify any prior OFFER. For example, in Turn 7, Dave offers a donkey as a gift suggestion, but in Turn 8, Ellen partially rejects the offer, indicating that she prefers something bigger.

At times, the last agent's OFFER will be incompatible with the current agent's internalized emergent. This will result in a *resolution action*: (1) accept the last OFFER, and modify the internal emergent; (2) construct an action which semantically resolves the contradiction; (3) *reject* the last agent's OFFER. This last option is dispreferred in human conversation, and often leads to argument or negotiation.

4. The emergent frame is updated.

The offer is provisionally entered into the collective emergent. Offers become increasingly stable as they are accepted and elaborated by other agents in later actions.

5.4 Metapragmatic Strategies and Power

In improvisational interaction, the features of the emergent frame are negotiated using *metapragmatic strategies*. This aspect of communication is "metapragmatic" because it reflexively comments on the pragmatic level of interaction [50]. For example, pragmatic features of an interaction include (among other features) the role relations of the participants, and the speech styles and interactional norms that are appropriate for conversation. A communication serves a metapragmatic function when it negotiates these pragmatic features. Such negotiation becomes particularly critical when the pragmatic features of the interaction are not specified in advance. Although this is often the case in human communication, this situation is exaggerated in improvised theater dialogue, since none of the pragmatic features of the encounter are pre-determined.

Metapragmatic strategies range from *relatively powerful* to *relatively weak* [42]. Both the response and the offer can employ distinct strategies. A powerful offer constrains the range of potential actions of the subsequent agent; a weak offer, in contrast, provides the next agent with a wide range of possible actions.

As an interaction proceeds, the interactional frame exerts downward causation, a collective form of metapragmatic power. In an improvised theater performance, as each minute of the drama proceeds, the dramatic frame is more broadly and more deeply defined. Thus, the frame gradually acquires more metapragmatic power (downward causation) of its own, and actors increasingly have to construct their turns so that they will work effectively within the active emergent frame. Consequently, the power vector of each offer must be considered to pass through the emergent frame, rather than passing directly from one agent to another. Note in particular the contrast with the reductionist atomism of most social simulations, which assume that all communications pass between pairs of agents, unmediated by any higher-level phenomena.

Studies of natural discourse have shown that rejections require the use of a more powerful strategy, whereas acceptances can employ a weaker strategy [44]. Studies have also shown that rejecting a powerful strategy requires an even more powerful strategy to be used in response. Variables that contribute to the interactional power of a strategy include:

- the primary implications of the offer (ranging from OTHER AGENT to SELF)
- whether or not the offer is targeted (TARGETED to a specific agent, or BROADCAST to all participating agents)
- whether or not the agent steps out of the current interactional frame (communication channel) to make the offer (IN-FRAME or OUT-OF-FRAME strategies)

The metapragmatic function of interaction is of a different order than pragmatic functions such as speech acts. Any given speech act can be formulated using any type of metapragmatic strategy. A directive can be other-directed and targeted—"John, please close that window." But it can also be formulated as self-directed and non-targeted—"I'm getting cold, maybe I'll put on a sweater" said as an indirect way of getting someone to close the window.

The most powerful strategies used in natural discourse correspond to the speech acts found in BDI-type ACL. All of the speech acts in these ACL are *explicit primary performatives*, and are denotationally explicit and targeted; negotiatory communications are distinct messages, and thus can be considered to pass through a separate interactional channel. However, in everyday human discourse, most interactional turns are composed at the opposite extreme of all three dimensions: humans display a marked preference for implicit, untargeted, and in-frame communications.

Thus, current ACL, although perhaps effective for applications such as Internet tools, are insufficient as models of natural discourse. Unfortunately for those interested in social simulation, ACL are lacking those characteristics most likely to result in collaborative emergence.

6 Conclusion

In many computational models of emergence, the endstate of the system is of primary interest. In collaborative emergence, the most interesting aspects of the simulation will be the incremental processes of emergence and downward causation.

Collaborative emergence models could be used in virtual reality applications, and in human interfaces where natural language communication is desirable, in spite of its ambiguities.

Improvisational models could also be used to generate narrative structures. These may or may not be similar to stories that are composed by a single author; for example, many of the narratives created by improvisational ensembles have a high degree of *local coherence* but very little *global coherence*. The result of such a simulation would be a representation of the narrative structures that are created, along with the remaining uncertainty, degrees of slippage, and unrealized offers.

Perhaps the most useful application of such collaborative emergence models will be in social simulation. It is notable that computational modeling has reproduced a longstanding sociological distinction, that between micro and macrosociology. Macrosociology models large social systems and stable structures like institutions, and thus roughly parallels reactive agent MAS; microsociology models small-group symbolic interaction and complex communication, and thus roughly parallels cognitive agent MAS. As Castelfranchi [7] noted, (macro) emergence in cognitive agents MAS remains unexplored.

Alife and MAS have had some success at modeling social species like ants and termites. Such species have the advantage that they are compatible with both atomistic theory, and with the assumptions built into the modeling tools—many agents, simple communications, and no downward causation. However, MAS have had difficulty ramping up to human social systems. Our explorations of collaborative emergence and of sociological theory suggest why this is the case.

I have suggested some ways to connect MAS more directly with sociological theories, particularly those that theorize symbolic interaction. As our models of social systems become more robust, we should expect that our theory and our models will begin to approach the complexity of natural social systems.

References

1. Agre, P. E. 1995. Computational research on interaction and agency. In P. E. Agre and S. J. Rosenschein (Eds.), *Computational theories of interaction and agency* . Cambridge: MIT Press.
2. Alexander, J. C., Giesen, B., Munch, R., and Smelser, N. J. eds. 1987. *The Micro-Macro Link* . Berkeley, CA: University of California Press.
3. Baas, N. A. 1994. Emergence, Hierarchies, and Hyperstructures. In *Artificial life III, SFI studies in the sciences of complexity, Proc. Vol. XVII*, edited by C. G. Langton. Reading, MA: Addison-Wesley, p. 515-537.
4. Bechtel, W., and Richardson, R. C. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.
5. Campbell, D. T. 1974. “Downward Causation” in Hierarchically Organised Biological Systems. In *Studies in the Philosophy of Biology: Reduction and Related Problems*, edited by In F. J. Ayala and T. Dobzhansky. Berkeley, CA: University of California Press, p. 179-186.
6. Cariani, P. 1991. Emergence and Artificial Life. In *Artificial Life II, SFI Studies in the* Sciences *of Complexity, Vol. X*, edited by C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen. Reading, MA: Addison-Wesley, p. 775-797.
7. Castelfranchi, C. 1998. Simulating with Cognitive Agents: The Importance of *Cognitive Emergence*. In *Multi-Agent Systems and Agent-Based Simulation*, edited by J. S. Sichman, R. Conte and N. Gilbert. Berlin: Springer, p. 26-44.
8. Chapman, D. 1987. Planning for Conjunctive Goals. *Artificial Intelligence* 32: 333-378.
9. Clark, H. H., and Brennan, S. E. 1991. Grounding in Communication. In *Perspectives on Socially Shared Cognition*, edited by L. B. Resnick, J. M. Levine, and S. D. Teasley. Washington, DC: American Psychological Association, p. 127-149.
10. Clark, A. 1997. *Being There: Putting Brain, Body, and World Together Again*. Cambridge: MIT Press.
11. Coleman, J. S. 1990. *Foundations of Social Theory*. Cambridge: Harvard University Press.
12. Collins, R. 1981. On the Microfoundations of Macrosociology. *American Journal of Sociology 86*: 984-1014.
13. Darley, V. 1994. Emergent Phenomena and Complexity. In *Artificial Life IV, Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, edited by R. A. Brooks and P. Maes. Cambridge: MIT Press, p. 411-416.
14. Davidson, D. 1970. Mental Events. In L. Foster and J. W. Swanson, *Experience and Theory*. Amherst, MA: University of Massachusetts Press, p. 79-101.
15. Durkheim, E. 1964. *The Rules of Sociological Method*. New York: The Free Press. Originally published in French in 1895.
16. Engelmore, R., and Morgan, T. eds. 1988. *Blackboard Systems* . Reading, MA: Addison-Wesley.
17. Epstein, J. M., and Axtell, R. 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. Cambridge, MA: MIT Press.

18. Faith, J. 1998. Why Gliders Don't Exist: Anti-Reductionism and Emergence. In *Artificial Life VI*, edited by C. Adami, R. K. Belew, H. Kitano, and C. Taylor. Cambridge: MIT Press, p. 389-392.
19. Fodor, J. A. 1989. Making Mind Matter More. *Philosophical Topics 17*: 59-79.
20. Giddens, A. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California Press.
21. Gilbert, N. 1995. Emergence in Social Simulations. In *Artificial Societies: The Computer Simulation of Social Life*, edited by N. Gilbert and R. Conte. London: University College London Press Ltd, p. 144-156.
22. Gilbert, N., and Conte, R. eds. 1995. *Artificial Societies: The Computer Simulation of Social Life*. London: UCL Press.
23. Grand, S., and Cliff, D. 1998. Creatures: Entertainment Software Agents with Artificial Life. *Autonomous Agents and Multi-Agent Systems*, 1: 39-57.
24. Hayes-Roth, B., Brownston, L., and van Gent, R. 1995. Multiagent Collaboration in Directed Improvisation. In *ICMAS-95*, edited by V. Lesser. Cambridge: MIT Press, p. 148-154.
25. Hutchins, E., and Hazlehurst, B. 1995. How to Invent a Lexicon: The Development of Shared Symbols in Interaction. In *Artificial Societies: The Computer Simulation of Social Life*, edited by N. Gilbert and R. Conte. London: University College London Press Ltd, , p. 157-189.
26. Jennings, N. R., Sycara, K., and Wooldridge, M. 1998. A Roadmap of Agent Research and Development. *Autonomous Agents and Multi-Agent Systems*, 1: 7-38.
27. Kelso, M. T., Weyhrauch, P., and Bates, J. 1992. *Dramatic Presence*. Technical report CMU-CS-92-195, Department of Computer Science, CMU.
28. Knorr-Cetina, K., and Cicourel, A. V. eds. 1981. *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies*. Boston: Routledge and Kegan Paul.
29. Kronenfeld, D. B., and Kaus, A. 1993. Starlings and Other Critters: Simulating Society. *Journal of Quantitative Anthropology*, *4*: 143-174.
30. Langton, C. G. ed. 1994. *Artificial Life III: Proceedings Volume XVII, Santa Fe Institute Studies in the Sciences of Complexity*. Reading, MA: Addison-Wesley.
31. Lesser, V. ed. 1995. *ICMAS-95: First International Conference on Multi-Agent Systems*. Cambridge: MIT Press.
32. Maruichi, T., Ichikawa, M., and Tokoro, M. 1990. Modeling Autonomous Agents and their Groups. In *Decentralized AI: Proceedings of the First European Workshop on Modeling Autonomous Agents in a Multi-Agent World*, edited by Y. Demazeau and J.-P. Muller. NY: Elsevier Science, p. 215-234.
33. Matusov, E. 1996. Intersubjectivity Without Agreement. *Mind, Culture, and Activity*, 3: 25-45.
34. Meyer, D. A., and Brown, T. A. 1998. Statistical Mechanics of Voting. *Physical Review Letters*, 81: 1718-1721.
35. Nicolis, G., & Prigogine, I. 1977. *Self-Organization in Nonequilibrium Systems*. New York: John Wiley & Sons.
36. Nishimura, S. I., and Ikegami, T. 1998. Emergence and Maintenance of Relationships Among Agents. In *Artificial Life VI*, edited by C. Adami, R. K. Belew, H. Kitano, and C. Taylor. Cambridge: MIT Press, p. 438-442.
37. Parsons, T. 1951. *The Social System*. Glencoe, IL: The Free Press.

38. Rao, A. S., and Georgeff, M. P. 1995. BDI Agents: From Theory to Practice. In *ICMAS-95*, edited by V. Lesser. Cambridge: MIT Press, p. 312-319.
39. Rawls, A. W. 1987. The Interaction Order Sui Generis: Goffman's Contribution to Social Theory. *Sociological Theory 5*: 136-149.
40. Sawyer, R. K. 1992. Improvisational Creativity: An Analysis of Jazz Performance. *Creativity Research Journal*, 5: 253-263.
41. Sawyer, R. K. 1995. Creativity as Mediated Action: A Comparison of Improvisational Performance and Product Creativity. *Mind, Culture, and Activity*, *2*: 172-191.
42. Sawyer, R. K. 1996. The Semiotics of Improvisation: The Pragmatics of Musical and Verbal Performance. *Semiotica*, *108*: 269-306.
43. Sawyer, R. K. 1997a. Improvisational Theater: An Ethnotheory of Conversational Practice. In *Creativity in Performance*, edited by R. K. Sawyer. Greenwich, CT: Ablex, p. 171-193.
44. Sawyer, R. K. 1997b. *Pretend Play as Improvisation: Conversation in the Preschool Classroom*. Norwood, NJ: Lawrence Erlbaum Associates.
45. Sawyer, R. K. 1999. The Emergence of Creativity. *Philosophical Psychology*, 12: 447-469.
46. Sawyer, R. K. 2000. *Bi-directional Causation: Unifying Reductionism and Emergence*. Paper presented to the Program in Philosophy, Neuroscience, and Psychology at Washington University in St. Louis, February 25.
47. Sawyer, R. K. 2000. *Creating Conversations: Improvisation in Everyday Discourse*. Cresskill, NJ: Hampton Press.
48. Sawyer, R. K. 2000. Durkheim's dilemma: Toward a sociology of emergence. Unpublished manuscript.
49. Sawyer, R. K. 2000. Nonreductive Individualism: Social Emergence Arguments from Philosophy. Unpublished manuscript.
50. Silverstein, M. 1993. Metapragmatic Discourse and Metapragmatic Function. In *Reflexive Language*, edited by J. A. Lucy. New York: Cambridge University Press, p. 33-58.
51. Sperry, R. W. 1986. Discussion: Macro- versus Micro-Determinism. *Philosophy of Science*, 53: 265-270.
52. Steels, L., and Kaplan, F. 1998. Stochasticity as a Source of Innovation in Language Games. In *Artificial Life VI*, edited by C. Adami, R. K. Belew, H. Kitano, and C. Taylor. Cambridge: MIT Press, p. 368-376.
53. Walker, A., and Wooldridge, M. 1995. Understanding the Emergence of Conventions in Multi-Agent Systems. In *ICMAS-95*, edited by V. Lesser. Cambridge: MIT Press, p. 384-389.
54. Winch, P. 1958. *The Idea of a Social Science*. Atlantic Highlands, NJ: Humanities Press International.

Modeling a Virtual Food Court Using DECAF*

Foster McGeary and Keith Decker

Computer and Information Sciences Department, University of Delaware
Newark, DE 19716
{mcgeary,decker}@cis.udel.edu

Abstract. Since voluntary organizational decisions are most likely made by self-interested parties, organizational decisions in an economic market are subject to explanation by (at least) two different fields of study: organizational theory and microeconomics. This paper describes initial work on economic modeling, including the modeling of voluntary organizational contracts, of a set of business entities set in the market context of the food court at an upscale suburban mall, the Virtual Food Court. DECAF is an multi-agent toolkit that provides this effort with a GUI to easily establish the control mechanisms needed to express exchanges between participants and the operating system features to perform the message sending and receiving that models the actual transactions. DECAF provides high-level language support and pre-built middle-agents for building multi-agent systems. This approach helps researchers or students get to the "interesting" parts of a simulation more quickly.

1 Introduction

Economic entities exhibit many of the characteristics associated with agents. First, each economic entity is (generally) viewed as a rational decision-maker attempting to maximize its own welfare. An economic entity can thus be thought of as a rational computational agent accomplishing a goal. Second, each economic entity/agent is concerned with and in control of its own welfare, and is thus an autonomous agent. Third, each agent is viewed as intelligent because it engages in purposeful activity expected to achieve goals or objectives. Having made these three observations, we use the DECAF agent toolkit [5] to model potential individual consumer and individual business behaviors explicitly, giving us a collection of economic agents which we expect to behave as received economic theory instructs us. Our agents have limited rationality (unlike traditional economic agents), can engage in certain specific well-defined behaviors, and have explicit representations of how to compute their own self interest. We then let the individual agents communicate and bargain with each other, where each agent only makes those transactions, choosen from among those available to be made with other autonomous agents, that maximize its own individual

* This work was supported by the National Science Foundation under grant IIS-9733004.

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 68–81, 2000.

best interests, as best it can. The aggregate behavior of such agents should provide the same results as received economic and organizational theories (as far as such theories rely primarily on economic behavior). Our modeling approach provides us with the ability to prohibit certain selected behaviors, and thus to identify those key individual behaviors which are essential to the production of a particular behavior.

VFC models diners, workers, and entrepreneurs. These economic entities are caricatures of the participants in transactions that take place within a simplified shopping mall food court. They exhibit suffcient behaviors to allow VFC to contain a labor market, markets for food service equipment, and markets for food products. For example, accepting a contract to perform labor and forming an organization are reciprocal events. Because both of these are voluntary actions, we believe it necessary to model and explain both sides of the transaction simultaneously. This is what we do in VFC, planning to extend our results to model organizational structures more complicated than a simple employment contract while still basing the analysis on the need for there to be reciprocally voluntary contracts. We expect that such models will be expanded to include aspects of governance [7] and non-economic social forces [4] as we explore the long term control and stability of such structures.

VFC exhibits behavior in a model of bounded rationality. Information is only available through reporting mechanisms or communications with other entities. Not all information is available to all decision makers. Limits are be placed on the amount of information that can be analyzed and the amount of time available to analyze it. These limits are a reasonable reflection of actual decision making, which is often bounded by time. DECAF provides a convenient mechanism for modeling economic activity as the actions of a set of agents. In particular, DECAF's model of software agency includes as primary features the need to support reasoning under deadlines, reasoning about alternatives, and reasoning about tradeoffs between multiple activities.

2 Our Principal Interests

Management of an arbitrarily large number of interacting agents and organizing them into workable units, while still recognizing and respecting their individual autonomy, is a significant problem. Since multi-agent system toolkits such as DECAF do not limit the number of agents, since the various agents can reside on different machines, and since the computational location of an agent is of no concern for modeling purposes, simple computational scalability does not loom as a significant problem. However, when these agents must interact in complex, changing ways, the scalable control and coordination of such systems becomes a pressing issue. One principal interest is, then, how to organize and at least partially control large dynamic multi-agent systems.

DECAF allows us the opportunity to create an artificial world with all the essential characteristics of the economic market place. With this tool, we can study and explain the "management" of numerous autonomous agents to simulate the economic activities which characterize the agent-selected activities that result from the mutually independent desires of each agent to maximize its own well being. To fully achieve such an explanation, we must further explain how autonomous agents communicate and interact with one another within constraints imposed by physical reality, to achieve organizations. We put quotes around the word "management" because, as Mintzberg would have it, there is mutual adjustment between agents. The only decisions are those of individuals, and thus organization is an observable result of the communication and actions of individual autonomous agents. Explaining how organizations come about out of large numbers of autonomous individuals, in a changing environment, with changing technologies and changing demands, is our principal long term goal. VFC serves as the testbed for those ideas as a means of explaining how businesses organize.

3 Overall Operation of the Virtual Food Court

As with the general economy, activity in VFC is consumer (Diner) driven. Each Diner has the goal of maximizing its own satisfaction from food purchases. Diners' preferences for food are known to the Diner and are revealed through requests for items and selections from among the available foods. Entrepreneurs act as vendors and attempt to maximize their incomes (the difference between sales revenue and total labor and preparation costs). Information on labor and food purchases is available through "governmental" bureaus that record and publish some (but not all) information related to labor use and to food consumption. As a result of the interaction of diners, workers, and entrepreneurs, restaurants are formed (organized) to serve the needs of the diners. The restraunts, in turn, serve the income desires of the entrepreneurs. Alternative forms of ownership and organization are expected to evolve in later versions of the VFC. Notions of market purchases and contracts are used in VFC. Market purchases are simple exchanges of money for food products or for one act of labor where each exchange is monetarily independent of any other such exchange. A contract would be any other arrangement, covering more than one instance or exchange.

A Typical Study. Our eventual goal in a typical study with VFC is to explain, step-by-step, some specific and particular economic phenomenon, say the equilibriation of a purely competitive market. We believe it is not sufficient to show simply that such phenomena can happen, but rather that is necessary to detail the conditions under which the studied phenomenon happen as a result of voluntary action. With competitive markets, we would need to show the conditions which need to exist so that entrepreneurs, operating in their own best interests, act in ways consistent with competitive markets. Models which require agents to act in other than their own best interests should not be demonstrable using the VFC. We believe that economic actions modeled on autonomous agents, using agents that are smart enough to collude or engage in self-serving anti-competitive actions, provides more information than models that simply assume the ideal conditions, assume unlimited capability on the part of the agents, and thus assume their conclusion.

A Simpler Question or Study. Transaction cost economics places a great deal of weight on the behavioral and engineering assumptions on which it is built. The behavioral emphasis stresses cost minimization and self interest. The engineering assumptions stress particular production functions (in VFC, these are the technologies and the parameters which differentiate them). In total, though, transaction cost economics deals with only simple cost minimization, and the contracts that occur are those needed to secure the necessary resources.

A simpler question to study is how the nature of the production functions, such as their "lumpiness," leads to one market form rather than to another. Similar questions or demonstrations, such as how the introduction of a technology with increasing returns to scale affects market structure, can be studied, including the provision of a trace of how the change comes about.

A more Complicated Question or Study. It is possible to enrich the knowledge base and behaviors of the participants (firms and workers) to more fully investigate organizational behavior. For example, explicit representations of how expectations are formed can be used to more fully demonstrate organizational evolution.

Organizational theory says that the organization behaves differently if it expects to repeat a transaction than it does if the transaction is not to be repeated. It could be the case that a transaction cannot be repeated if the firm behaves one way rather than another. There are a large number of such situations (chicken and egg questions) in organizational theory (and in transaction cost economics, too). We would like to investigate how the organization makes the determination of whether it expects to repeat a transaction. Similarly, there are organizational theories that rely on more than bounded economic self-interest, including the influence of social networks and patterns of information flow (e.g. [12]).

Bounded Rationality. Economic activities occur through time. The duration between the creation and expiration of an economic opportunity affects each agent's ability to exploit it. A model of microeconomic decisionmaking is more realistic if it can trace the making of decisions with respect to the timing of decisionmaking and the identification of the information needs of successful decisionmakers.

Knowledge-based decisionmaking is bounded by the time and limited bandwidth available for communication. It is unrealistic to expect decisionmaking to be based on full information. VFC allows explicit modeling of that boundedness, and allows the modeling to take place at the level of the individual worker or production unit. We believe models that do not adequately represent the boundedness of decisionmaking do not properly explain decisionmaking or the economic and organizational evolution that comes from such decisions.

Because we can limit agent memory and performance, and thus create agents with different memory and performance capabilities, we can create competitions between those who require more information, and therefore respond to economic stimuli more slowly, and those who require less information, and thus respond more quickly. Careful construction of those agents and agent behaviors should allow the study of how boundedness affects microeconomic decisionmaking.

4 Overview of DECAF's Internal Agent Architecture

DECAF (Distributed, Environment-Centered Agent Framework) is a toolkit which allows a well-defined software engineering approach to building multi-agent systems. The toolkit provides a stable platform to design, rapidly develop, and execute intelligent agents to achieve solutions in complex software systems. DECAF provides the necessary architectural services of a large-grained intelligent agent: communication, planning, scheduling, execution monitoring, and coordination. This is essentially, the internal "operating system" of a software agent, to which application programmers have strictly limited access.

The goals of the architecture are to develop a modular platform suitable for our research activities, allow for rapid development of third-party domain agents, and provide a means to quickly develop complete multi-agent solutions using combinations of domain-specific agents and standard middle-agents [1]. DECAF distinguishes itself from many other agent toolkits by shifting the focus away from the underlying components of agent building such as socket creation, message formatting, and the details of agent communication. Instead, DECAF provides an environment that allows the basic building block of agent programming to be an agent action, or a pre-specified subtask (collection of agent actions). These building blocks are then chained together by the DECAF planner. This paradigm differs from most of the well known agent toolkits, which instead use the API approach to agent construction (e.g., [8]). Functionally, DECAF is based on RETSINA [3] and TAEMS[2,13].

Although a traditional HTN planning component is in development, currently the control or programming of DECAF agents is provided via a picture-based GUI called the *Plan-Editor*. The Plan-Editor can also be used to construct shared task network libraries for common multi-agent protocols. The GUI can also be used to visually examine automatically-constructed plans for either DECAF or TAEMS.

Figure 1 represents the high level structure of the DECAF architecture. Structures inside the heavy black line are internal to the architecture and the items outside the line are user-written or provided from some other outside source (such as incoming messages).

As shown in Figure 1, there are five internal execution modules (square boxes) in the current DECAF implementation, and seven associated data structure queues (rounded boxes). DECAF is multi-threaded, and thus *all modules execute concurrently, and continuously* (except for agent initialization).

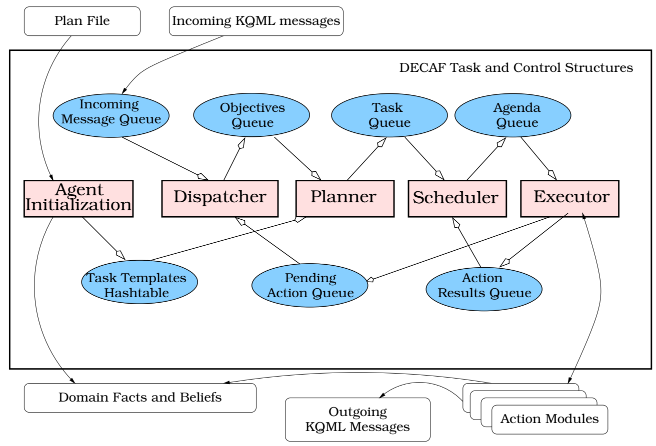


Fig. 1. DECAF Architecture Overview

Agent Initialization. When an agent is started, the *Agent Initialization* module will read a plan file containing basic action definitions and pre-defined task network reductions. Each task reduction specified in the plan file will be added to the *Task Templates Hashtable* (plan library), indexed by the particular goals that the reduction achieves.

The agent initialization process also attempts to achieve a *Startup* goal. The Startup tasks of a particular agent might, for example, build any domain data/knowledgebases needed. Startup tasks may assert certain continuous maintenance goals or other initial achievement goals for the agent. The last thing the Agent Initialization Module does is register with an Agent Name Server and set up all socket and network communication.

Dispatcher. The Dispatcher waits for incoming KQML messages which will be placed on the *Incoming Message Queue*. An incoming message contains a KQML (or FIPA) *performative* and its associated information. An incoming message can result in one of three actions by the dispatcher. First, the message may be a part of an ongoing conversation. In this case the dispatcher will find the corresponding action in the *Pending Action Queue* and provide the incoming message as an enabling provision (see "Parameters and Provisions", below). Second, a message may indicate that it is part of a new conversation. If so, a new *objective* is created (equivalent to the BDI "desires" concept[9]) and placed on the *Objectives Queue* for the Planner. An agent typically has many active objectives, not all of which may be achievable with an agent's limited time and resources. The last thing the Dispatcher is responsible for is the handling of error message replies to malformed messages.

Planner. The Planner monitors the Objectives Queue and plans for new goals, based on the action and task network specifications stored in the Plan Library. A copy of the instantiated plan, in the form of an HTN corresponding to that goal is placed in the *Task Queue* area, along with a unique identifier and any provisions that were passed to the agent via the incoming message. The Task Queue at any given moment will contain the instantiated plans/task structures (including all actions and subgoals) that should be completed in response to all incoming requests and any local maintenance or achievement goals. A graphical plan-editor GUI allows the construction of static HTN planning task structures.

Scheduler. The *Scheduler* waits until the Task Queue is non-empty. The purpose of the Scheduler is to determine which actions *can* be executed now, which *should* be executed now, and in what order.

For DECAF, the traditional notion of BDI "intentions" as a representation of a currently chosen course of action is partitioned into three deliberative reasoning levels: planning, scheduling, and execution monitoring. This is done for the same reasons given by Rao [9]—that of balancing reconsideration of activities in a dynamic, real-time environment with taking action [10]. Rather than taking the formal BDI model literally, we develop the deliberative components based on the practical work on robotics models [11]. Each level has a much tighter focus, and can react more quickly to external environment dynamics than the level above it. Most authors make practical arguments for this architectural construction, as opposed to the philosophical underpinnings of BDI, although roboticists often point out the multiple feedback mechanisms in natural nervous systems.

Once an action from the Task Queue has been selected and scheduled for execution, it is placed on the *Agenda Queue*. In a very simplistic Scheduler, the order might be first-come-first-served (FCFS). Recent work has resulted in the development of a sophisticated *design-to-criteria* action scheduler [13] that efficiently reasons about action duration, cost, result quality, and other utility function characteristic trade-offs. This scheduler is also available for use in DECAF agents.

Executor. The *Executor* is set into operation when the Agenda Queue is non-empty. Once an action is placed on the queue the Executor immediately places the task into execution. When the action completes it signals a specific action outcome, and the result is placed on the *Action Result Queue*. The framework waits for results and then distributes the result to downstream actions that may be waiting in the Task Queue. Once this is accomplished the Executor examines the Agenda queue to see if there is further work to be done.

5 Overview of the Virtual Food Court

VFC produces a series of arms-length voluntary exchanges of money for labor or the outputs of production. Capital products, restauraunt supplies, and food products for final consumption are the outputs of production. Technologies define the combinations of factors required to produce a particular product. The factors of production are capital products, labor, and, generally, other products. The initial configuration of VFC is shown in Figure 2 as a collection of boxes and lines. Lines represent the initial communications paths built into the Java code that constitutes each agent. Arrowheads indicate the direction of the initial message. To bootstrap the connection problem, agents know of the Matchmaker to register their existence with it. Workers and Restaurants know of the existence of the Government because they report information to it.

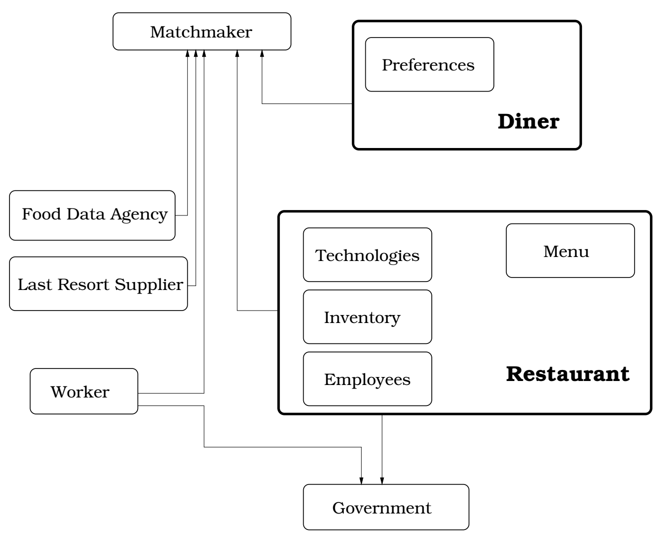


Fig. 2. Virtual Food Court Architecture

Boxes inside the Diner and Restaurant indicate lists associated with each instantiation of a diner or a restaurant. The lists represent that particular agent's knowledge, assets, and the values for parameter-driven behaviors. In general, no two diners have the same preferences, and no two restaurants have the same technologies, inventories, employees, or menu.

Matchmaker. Matchmaker is a DECAF facility that operates as a well-known referral service [1]. The standard term for this is "yellow-pages" because referrals are made on the basis of declared values submitted by agents that choose to "advertise" with the Matchmaker. Agents that wish to supply services (except the Government, which considers itself well-known) advertise with the Matchmaker. Diners do not register with the Matchmaker, as they are users of services and do not need to advertise. There is within DECAF the ability for agents to "subscribe" to advertisements as well as to advertise. Use of this feature allows Diners to receive notices of Restaurant openings.

Diner. The Diner is the key agent to triggering activity. The Diner is sent an initial message by the investigator containing the number of meals to consume. The Diner begins an eat-select loop that terminates only after the required number of meals have been consumed.[1]

The Diner's Initial Condition. Diners are instantiated with no specific knowledge of the world other than their own name and how to contact the Matchmaker. Each Diner reads a list of items that are its preferred foods and its current preference for each. Currently all Diners have the same code, but can differ in their initial preferences. As the Diner is more fully developed, the addition of various alternative methods of choosing restaurants and forming preferences can be contemplated.

How a Diner Behaves. Figure 2 shows only the initial conditions for the VFC, with the lines indicating potential message flows. The Diner asks the Matchmaker for the names of all restaurants and selects one at which to procure a meal. This action provides the Diner with the information to start a conversation with any of the Restaurants that have advertised with the Matchmaker and which meet any other criteria of the Diner. The Diner "enters" the selected Restaurant by sending an appropriate message. It receives a Menu from the Waiter, and performs a calculation that creates a ranked list of items (based on preference and price). With its ranked preferences, the Diner then proceeds, through a series of messages and replies, to conduct a typical restaurant conversation. After a meal, the Diner updates its priorities by decreasing those of items it has received and ncreasing those of items it has not received. If the assigned number of meals have not been consumed, the Diner chooses a new restaurant at which to procure the next meal.

Restaurant. The Restaurant begins by reading files containing its initial inventory, its initial menu, and its initial stock of technologies. The restaurant then asks the Matchmaker for the name and address of the supplier of last resort and for the name and address of a Food Data Agency. The Restaurant then hires a waiter and advertises itself with the Matchmaker. When a Diner enters the Restaurant, the Restaurant assigns a Waiter to the Diner, and sends the Waiter the current menu, inventory, and name and address of the Diner. Once the Waiter delivers the order, the Restaurant tries to fulfill the order from inventory, and then restock inventory. If the waiter quits, the Restaurant hires a new one. If the waiter does not quit, the Restaurant waits for the next Diner to enter.

[1] We like to call this the "Dining Economists" problem.

How a Restaurant Prepares a Meal. The current paradigm is one of only serving items already prepared and in inventory. If an item is on the Menu, a Diner may ask a Waiter for it. The Waiter has a copy of the inventory at the time the Diner entered, so it may accept an order for an item that has disappeared from inventory before the order is serviced. Ordered items are deleted from inventory when "sold" to the Diner. After each order is processed, a message is sent to the Waiter listing which items were available, and another is sent to the Government as Delivered items.

How a Restaurant Restocks. Restaurants produce to satisfy inventory requirements, subject to minimum order quantities. That is, for each item on the menu there is a desired number of the items to have in inventory. After each diner's order is fulfilled, the restaurant compares its then-current inventory to the desired levels. If an item is below its desired level by at least the minimum order quantity, then the restaurant attempts to replenish its inventory either by making the item (if has the technology and the resources) or by purchasing the items from a supplier.

Initially, the only supplier is the supplier of last resort, which can supply any item in any quantity - hardly a model of reality. However, by having the supplier of last resort charge "too much" for certain items, we can create the economic incentives needed for suppliers to compete with it. Since we want all this activity to be driven by individual agents, we add business entities that are suppliers of restaurant supplies rather than restaurants. The Restaurant at the end of the supply chain can choose from more than one supplier, or can choose to make an item itself.

How a Restaurant Learns a New Technology. The potential for many make-or-buy decisions exists within a restaurant or other supplier. It must decide what items it will offer, which it can only decide after it knows what items there are and for what items there is a market. In addition, the supplier needs some notion of how profitable an item will be. This turns out to be both simpler and more complicated than it may appear. The process is simpler in a fully functioning VFC because there is market data that allow a supplier to evaluate the decisions. However, before the market is fully developed, early participants must acquire information in a manner that is best described as speculative.

Each restaurant/supplier has a parametrically driven algorithm for producing items. A Technology consists of a list of the parameters to the algorithm and a name for the output. The supplier of last resort is the initial supplier of Technologies. Technology transfer consists in the supplier of last resort selling the technology, which is the simple act of sending the list of parameters in response to a message request. The requester then has all the knowledge it needs to produce the item, although it must still acquire the inputs to apply the knowledge. As each Technology may require further items for which the restaurant does not have the Technology or the material in stock, several rounds of make-or-buy decisions may need to take place before anything is actually produced.

Government. VFC contains two governmental functions. One is a data reporting and analysis function (the Bureau of Food Statistics, or BFS), the other is a simple fact distribution service (the Food Data Agency, or FDA). The FDA is a repository of nutritional information on food products. The BFS accepts information from reporters and provides information to requesters.

FDA is intended to be used for several purposes further along the development path. It is our intention to trace all the information used by the various agents. Therefore, we will want to trace how diners knew to ask for donuts at restaurants, and not to ask for doorknobs. The FDA is the arbiter declaring things to be foodstuffs, and diners would need to verify with the FDA that an item they had heard about was a food.

BFS accepts reports that are paired lists of food items and number of servings. Three categories of food reports are accepted: Requested, Ordered, and Delivered. Requested items are those about which diners inquired, and are without regard to whether the item was even carried at the restaurant at which the request was made. Ordered items are those which the diner ordered at a restaurant that carries the item. Delivered items are those ordered items that the restaurant actually prepared.

Use of the data is two-fold. First, it allows the menu-design market to exploit the difference between what diners inquire about (Requested) and what is available (Ordered or Delivered) to revise menus or to add restaurants filling gaps in diner satisfaction. Second, the difference between Ordered and Delivered indicates a (market-adressable) problem with production.

Last Resort Supplier. The supplier of last resort provides whatever raw materials or manufactured goods are requested of it. Researcher control of the inputs to the diner and supplier of last resort should ensure that any product requested has a production function.

Markets for capital goods and for materials are closely related: both are manufactured goods, but materials depreciate much faster than do capital items. The essential difference is that use of a capital item can be transformed into a service material through the simple act of leasing the capital for a fraction of its useful life.

All non-labor factors of production are themselves outputs of production processes. The specification of the inputs to produce a particular output is done using a production function, which is termed a technology in this paper. With this recursive definition, of course, the process must come to some basic factors that are not themselves produced (land, air, fire, water).

Worker. Labor is modeled as a set of individual workers, each of which comes to the labor market with a potentially unique set of capabilities. Each worker is constructed as a separate agent, responsible for its own welfare, principally to be able to investigate individual behavior. Also, varying levels of labor skill is not presently implemented. There is only one worker model. The only skill or function this worker model has is that of selling, wherein the worker engages in a fairly simple exchange with the customer to establish and deliver a food order. The worker passes the order, via a message, back to the employer for fulfillment. Once the employer completes the order the worker presents the finished order to the customer. Workers not under contract become part of a central labor queue. From the central queue, workers can be recruited/assigned to any task. Workers may contract to work for an individual entrepreneur, and such contracting could be a major area of study with VFC.

Diners and Brokers. In VFC, the broker function (Figure 3) is viewed as one where an agent (Broker) is hired to perform a procurement function on behalf of a client. The Broker is expected to perform the procurement operation solely in the best interest of the client. To accomplish this, the Broker is given a list of items from the client and then performs a fixed sequence of events. First, the Broker asks the Matchmaker for a list of all "wholesale" suppliers registered there. Second, it send messages to each supplier asking for that supplier's menu (or "catalog"). Third, it processes each received menu (requests that timeout are not processed), looking for the least costly supplier to provide each item on the client's request list. Fourth, it orders the items from the least cost supplier; fifth it awaits delivery of the items; and sixth it sends a message to the client noting the items found and their cost.

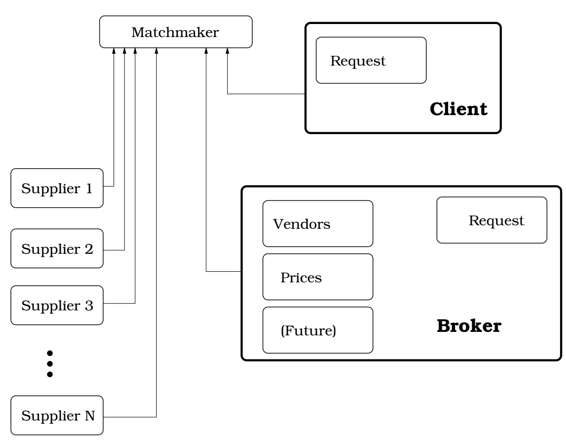


Fig. 3. Virtual Food Court Broker

Further, note the similarity in how a Broker would deal with a Restaurant or a supplier of pots and pans. All the Broker needs to know is what items the supplier supplies and how much they cost. In the future, quality issues will be addressed, as well as supply contracts.

6 A Specific Partial Example

Figure 4 shows the two conversations required for a VFC restaurant to hire a waiter. The first conversation is quite simple: the prospective waiter ("Emp') places an advertisement with the Matchmaker ("MM") asserting that the worker agent has waiter skills.

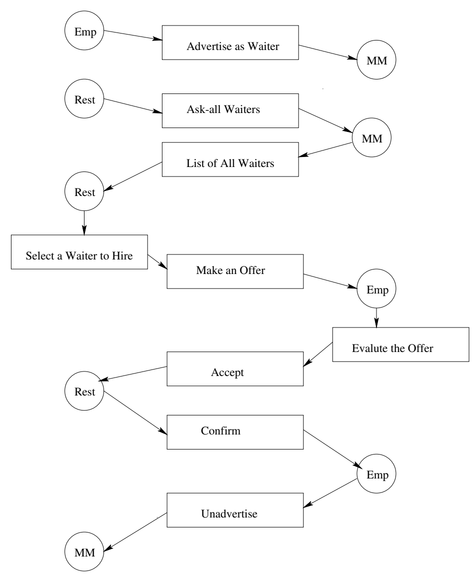


Fig. 4. Hiring Conversation

For the restaurant to hire an employee, a more complicated conversation is required. The restaurant must discover what agents are available to function as waiters. For the conversation in Figure 4, we have the restaurant ("Rest") issue a KQML "ask-all" message to the Matchmaker. The matchmaker responds with a list of agents that have advertised as having waiter skills. Upon receipt of the list, the restaurant decides which of the waiters it would like to hire, and makes an offer to the waiter through a KQML message sent to the prospective employee. The worker agent receives the employment offer from the restaurant, and decides whether to accept it or not. We have shown the conversation with an acceptance, which the worker sends to the restaurant. The restaurant confirms the acceptance, and the employee sends an unadvertise message to the Matchmaker to make itself unavailable to other employers.

7 Conclusion and Future Work

Our development proceeds along two lines. In one line, we are developing the technical production functions and agent-level knowledge storage requirements and mechanisms of the agents. In the other line, we are improving the behavioral repetoire and communication skills of the individual agents. We are encouraged by the speed with which even simple collections of agents can collect and distribute knowledge, so our early emphasis will be on increasing the number of agents at work, particularly on the supply side. Our primary behavioral side issues currently are modeling how a business decides what products to produce and how to form prices. The primary technical side issues are in building a set of production functions (technologies) that reasonably express likely U-shaped short-run production functions for restaurants and manufacturers. As these issues are addressed, we expect the VFC to grow in size and in complexity of exhibited behavior.

References

1. K. S. Decker, K. Sycara, and M. Williamson. Middle-agents for the internet. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, pages 578–583, Nagoya, Japan, August 1997. 72, 75
2. Keith S. Decker and Victor R. Lesser. Quantitative modeling of complex computational task environments. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 217–224, Washington, July 1993. 72
3. Keith S. Decker and Katia Sycara. Intelligent adaptive information agents. *Journal of Intelligent Information Systems*, 9(3):239–260, 1997. 72
4. Elihu M. Gerson. On 'quality of life'. *American Sociological Review*, 41:793–806, 1976. 69
5. J. Graham and K. S. Decker. Towards a distributed, environment-centered agent framework. In *Proceedings of the 1999 Intl. Workshop on Agent Theories, Architectures, and Languages (ATAL-99)*, pages 162–175, 1999. 68
6. H. Mintzberg. *The Structuring of Organizations*. The Free Press, New York, 1979.
7. Charles Perrow. *Complex Organizations*. Random House, New York, 1986. 69
8. Charles J. Petrie. Agent-based engineering, the web, and intelligence. *IEEE Expert*, December 1996. 72
9. A. S. Rao and M. P. Georgeff. BDI agents: From theory to practice. In *Proceedings of the First International Conference on Multi-Agent Systems*, pages 312–319, San Francisco, June 1995. AAAI Press. 73, 74
10. Stuart Russell and Eric Wefald. *Do the Right Thing: Studies in Limited Rationality*. MIT Press, Cambridge, MA, 1991. 74
11. R. Simmons, R. Goodwin, K. Haigh, S. Koenig, and J O'Sullivan. A layered architecture for office delivery robots. In *Proceedings of the 1st Intl. Conf. on Autonomous Agents*, pages 245–252, Marina del Rey, February 1997. 74
12. Arthur L. Stinchcombe. *Information and Organizations*. University of California Press, Berkeley, CA, 1990. 71
13. T. Wagner, A. Garvey, and V. Lesser. Complex goal criteria and its application in design-to-criteria scheduling. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, Providence, July 1997. 72, 74
14. Oliver E. Williamson. *Markets and Hierarchies: Analysis and Antitrust Implications*. The Free Press, New York, 1975.
15. Oliver E. Williamson. *The Mechanisms of Governance*. Oxford University Press, New York, 1996.

How to Combine Reactivity and Anticipation: The Case of Conflicts Resolution in a Simulated Road Traffic

Sameh El hadouaj[1,2], Alexis Drogoul[2], and Stéphane Espié[1]

[1] INRETS
2, avenue Malleret Joinville, 94114 Arcueilcedex, France
{espie,hadouaj}@inrets.fr
[2]LIP6 - OASIS/MIRIAD, UPMC/CNRS UMR 7606
Boîte 169, 4 Place Jussieu 75232 Paris CEDEX 05 France
Alexis.Drogoul@lip6.fr

Abstract. In this article we present the method used to solve the conflicts that can happen between agents that represent simulated drivers in a simulated road traffic. This work is part of the ARCHISIM project, which aims at both simulating a realistic traffic evolution and making the behaviour of the simulated drivers credible for a human driver placed in a driving simulator. After having categorized the types of conflicts that can happen, and the constraints that determine the choice of a solving method, we propose a method that combines reactivity and anticipation. This method is based on the works of driving psychologists who work in the INRETS institute. We offer an exprimental validation of this method with respect to real data and discuss its advantages in the perspective of largest applications.

1 Introduction

Multi-agent simulation models are a tool which becomes more and more important for the analysis and the understanding of complex phenomena [23], mainly in human organizations (see [26]). In this article we address the particular case of a realistic road traffic simulation. This simulation is integrated in the INRETS ARCHISIM project which was created jointly by the Analysis and Regulation Department (DART) and by the driving psychology laboratory (LPC) in 1992 [16]. It has been the object of a collaboration led with the laboratory of computer science of Paris 6 (LIP6), since 1997, through a master training and a PhD thesis.

The ARCHISIM project aims at producing the global phenomenon of a road traffic by reproducing the behaviour of each driver participating to the simulation. In this model a driver is considered as an autonomous agent whose behaviour is based on the psychological studies led by the LPC [25]. These studies provide a model describing the driver's decision activity with respect his environment. The traffic produced by ARCHISIM results from the interaction of the behaviour of each agent with the

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 82-96, 2000.

regulation, the infrastructure of the road and the other users. The realism of the simulation is then a direct function of the quality of the behavioural model.

In this paper, we describe the process that we have adopted to produce a simulated traffic that could be interpreted as “realistic” by a road user. We are particularly interested in drawing a parallel between the problems that are raised by this type of simulation and the problems met in many multiagent systems, such as the choice of a conflict resolution method. The plan of the paper is as follows: first, in section 2, we describe the ARCHISIM project. Second, we introduce in section 3 the problems encountered by the designer of such a simulation, especially in the resolution of the conflicts that occur between the agents. Third, we give an overall view of the traditional methods used in DAI and their drawbacks. We then propose in section 4 a method that gracefully combines reactivity and anticipation. This method is validated experimentally in section, and we end by an overall view of the perspectives of research in this field

2 ARCHISIM

New traffic regulation possibilities are usually introduced to make the traffic surer and more comfortable. The simulation models allow to test and to evaluate these equipments without resorting to life-size tests, which are expensive and difficult to achieve. Therefore, the simulation models constitute a tool that is more and more used for traffic forecast and management, for example by local collectivities. These models can be categorised in two families:

- Macroscopic models: they describe road traffic dynamics in a global way by describing its flow, its density and its average speed. The best known in France are: *Simaut* [19], *Strada* [6] or *Meta* [20].
- Microscopic Models: they represent traffic in more details. These models deal with individual vehicles and try to focus on the interactions that occur between vehicles. They manipulate variables like the inter-vehicle time, the vehicle speed and its acceleration, and so on. In this category of models, we can mention *Casimir* [8], *Severe* [3], *Mixic* [2], or *Pharos* [21].

The analysis of a social organization such as road traffic could not be done without taking the interactions between its members into account [18]. By focusing precisely on the interactions, the microscopic models allow us to produce more realistic traffic situations than those produced by macroscopic models. However, both macroscopic and microscopic models have generally the goal to reproduce traffic flow laws identified by measure campaigns and are usually designed to solve a well-defined problem. For example, *Casimir* is designed to compare working strategies of isolated crossroads traffic lights. *Mixic* allows the study of the AICC (Autonomous Intelligent Cruse Control) impact on a number of consecutive crossroad sections.

The goal of ARCHISIM model is not only to produce a realistic traffic, but also to simulate a traffic composed of vehicles that behave individually in a credible way. But the "realism" with which the behaviour is simulated strongly depends on the quality of their interactions. Therefore, the choice of a multi-agent architecture gives us the possibility to model in a detailed way the interactions between the drivers, as well as their behaviour. The traffic flow laws obtained are then a consequence of the interactions of the drivers' behaviour with the regulation, the infrastructure of the road and the behaviour of the other users.

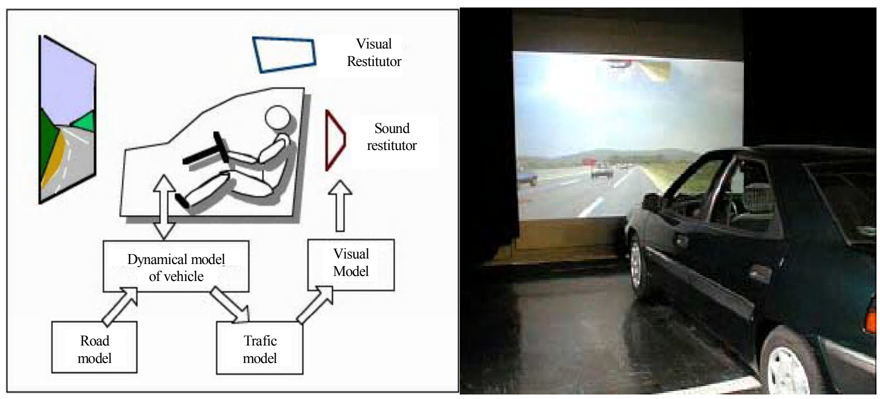


Fig. 1. Schema and view of the simulator

The modelisation of the behaviour of an artificial driver is based on the experimental driving psychology studies led by the LPC. This work makes use of a conceptual model defined by a set of rules which describe the driver decisional activities. The quality of this modelisation is a key element to the realism of the simulated road traffic situations (see figure 2).

Equally, the software architecture of the ARCHISIM model allows a simulator to participate into the simulation and thus a human driver to see and to interact with several traffic situations (see figure 1). The goal is to allow the test of "virtual" infrastructure or new equipments (driving aids, etc.) on "real" drivers, so that we could study their reactions. Consequently, the traffic situations produced by the model must give to a human operator the illusion to be in a real traffic situation, which represents a considerable challenge: the agents have to be credible and should not create unrealistic traffic situations. Eventually, ARCHISIM will then be able to [17]:

- test and refine models of behaviour;
- test man-machine interfaces on the simulator, check their compatibility with the driving task and assess their impact on the system;
- test the sensitivity of the system to various scenarios (such as, for example, the percentage of cars equipped with driving aids).

Fig. 2. An example of a simulated traffic (agents: infrastructure, vehicles, etc.)

3 Description of the Existing Multi-Agent System

The distributed nature of traffic, the absence of any central control and the importance of behaviours in the modelisation of the driving task have naturally led us to choose a multi-agent approach in order to model this phenomenon.

In ARCHISIM, we consider that a given traffic situation is the result of interactions between heterogeneous agents: vehicles, infrastructure, road lights controller, etc. (see figure 2). Each agent is an autonomous entity, possesses its own knowledge, goal and strategy in order to perform its various tasks and to solve any conflicts which might arise with other agents. The infrastructure "transmits" information to road users and is for this reason considered as an agent. However, we note that this agent is particular because of its singleness and static status in the system. It induces behaviours through its geometric dimensions (visibility distance, lane width, etc.) without being affected by other agent behaviours.

An agent updates its knowledge by exchanging data with its environment. An important problem is to provide each agent with the informations that could change its behaviour. This leads to determine the sets of neighbouring agents that are situated in its area of perception and that change at each step of time. The first solution we could think of is that each agent sends to the others a message containing its set of observable information, namely its position on the road. Then, each receiver decides according to this position whether the agent sender is in its scope of perception or not. However, with respect to the great number of agents involved in a traffic simulation and so the great amount of communications, we can see that such a solution can not maintain an acceptable execution time. These inter-agents communications have thus been replaced by the design of a “central” process which plays the role of a "vision server". This process is the only one provided with an overall knowledge of the road network status, but it has no access to the local knowledge of agents. For ARCHISIM, a simulation consists in starting up the "central" process. Each agent present in the simulation then sends a message to this "central" process that consists of its current status (position, speed, etc.) as well as a request concerning the elements that are in its area of perception. After having received all the messages from all the agents that participate to the simulation, the "central" process computes for each agent the set of elements that are included in its perception area and sends it to each agent.

4 Main Difficulties Related to Traffic Simulation

4.1 Interaction of a Set of Agents

In our model, the main agents are the drivers. The quality of the simulation mainly depends on them. In ARCHISIM, each driver is an autonomous entity that possesses its own behaviour, its own goal that consists in following a given itinerary and its own knowledge described by a partial representation of its environment. We assume that the driver and the vehicle represent a single entity. We do not try to model the interactions between those two elements. A driver undergoes a set of interactions that are described by constraints imposed by the infrastructure, the regulation and the other road users. On one hand, these interactions usually represent a source of conflicts, and on the other hand, the resolution of conflicts needs other interactions. We note that in this system as in most collective systems, the notions of interactions and conflicts cannot be dissociated. These conflicts are, at the same time, the cause and the effect of interactions. Therefore, having a realistic simulation requires the use of a conflict resolution method that can be relevant in the context imposed to us by this simulation, which means complying with the models proposed by the psychologists ("credibility") and ensuring real time operations.

4.2 Nature of Conflicts

In the context of road traffic we can distinguish several types of conflicts: resource conflicts, goal conflicts and commitment conflicts. Each of them is described in more details in the paragraphs below.

Resource Conflicts. The road network is defined by a set of roads and crossroads. A road is characterized by its infrastructure (its width, its lane number, etc.). The crossroads are the place of intersections of roads. A network can also be defined as a set of sections, where a section represents a part of a road and is characterized by an infrastructure and a capacity. The capacity represents the number of vehicles on this section, which have a reasonable chance to flow out during a time interval of reference [8]. Conflicts appear when the number of vehicles willing to run on the section becomes larger than its capacity. The traffic becomes dense and conflicts arise around the traffic space. These conflicts are qualified as resource conflicts. In fact, if the space were larger, or if the number of vehicles were smaller, the conflicts would disappear. The cause of the conflicts is thus the resources available to each vehicle. In traffic situations, we distinguish two kinds of resource conflicts:

- The lane insertion conflicts, that appear for example in the case of a closing lane. A vehicle that detects that the lane on which it travels will be closed should find a gap in an adjacent lane in order to enter it. If the traffic is dense on the adjacent lane in which the vehicle would like to go, conflicts could appear between this vehicle and those on the adjacent lane. The same type of conflict appears in the case of a highway with an access road, between the vehicles on the highway and the others which are on the access road. Likewise, when a vehicle decides to change lane and that the density of the lane on which it chooses to go is very important.

– The intersection conflicts generally appear at crossroads between vehicles coming from different roads and taking different directions that intersect. These vehicles should organize themselves and share the crossroads space so that each of them could take the direction it has already chosen and avoid collisions. The problem would not arise if there were a single vehicle or if the space were large enough for all the vehicles to flow at the same time.

Goal Conflicts. This type of conflicts appears even if the traffic on the section is not dense. They are caused by the simple presence of another vehicle. Therefore, a driver could be in conflict with another preceding vehicle, if this one runs with a speed smaller than the speed which the driver wants to reach. This type of conflict could be defined as a goal conflict. At a global level, the local objectives of each driver could not be merged in a single goal. The satisfaction of one's goal implies the dissatisfaction of the other's goal.

Commitment Conflicts. The driver could also be in conflict with a vehicle running ahead, which switches on its indicator at instant t and switches it off at instant t+1. The fact that the vehicle switches on its indicator implies that it is ready to leave the lane. The vehicle that is behind will take decisions on this basis, which it will cancel at instant t+1. This type of conflicts could be qualified as commitment conflicts. The promise made by the vehicle that switched on its indicator is not kept. All of the anticipations made by the other agents on this basis have to be canceled.

Some conflicts are predictable and can be avoided. In reality, they are usually solved by the use of constraints that are imposed by the infrastructure and the regulation. In fact, the Highway Code and the road equipments such as road lights and traffic signs impose hard constraints to the driver when he is driving on the road. Likewise, the highway was designed in a way that the flows that move in opposite directions are separated. This allows avoiding collisions that might occur between these two flows. These constraints are introduced into our model. They consist in a set of individual behaviour rules that aim to avoid collision between the two flows. In a multi-agent context, this resolution method is called coordination by regulation or social norms (see for example [9] who works on the emergence of social norms in an agent population, and shows the interest of introducing such constraints in a multi-agent system).

However, other conflicts exist that cannot be solved by this method. Sometimes, the application of the regulation leads to blocked situations. For example, in the case of a cross-shaped crossroad, on which a vehicle is placed at the end of each of the four roads: if drivers apply the left priority rule, none of them would decide to enter the intersection and all of them would be blocked. In practice, the driver uses his perception of the environment in order to make his decisions and to solve such situations. The driver agents should do so in order to have an individual and a collective behaviour that is realistic. Just like in reality, they should find a solution that satisfies most of the constraints imposed by the regulation, the infrastructure and the road traffic, without perhaps respecting all of them. In this article we are mainly interested in this type of conflicts, and in the methods allowing solving them.

5 Constraints for Conflicts Resolution

The psychological studies of driving led by the LPC [25], which inspired us during the design of the behavioural model, underlines the importance of the anticipation for the driver. In fact, the driver takes the forecasting duration of the interaction into account when reacting to it. He does not, simply, react to instantaneous variations of the data describing it. So if he predicts that the duration of the interaction will not be too long, he can decide to adapt and stay on his lane. Else he can try to eliminate this interaction. Therefore, the adopted method for conflict resolution should allow the driver agents to anticipate the traffic evolution, so that they could have a credible behaviour.

On the other hand, because of the dynamical character of the traffic, it is quite difficult to predict it accurately. For this reason, we need a quick and adaptive conflict resolution method in order to allow the simulated driver to react in real time, i.e. as quickly as a real driver. Given the constraints imposed upon the agent driver (mobility, movement, etc.), we are in fact close to the design of reactive robots, for example those participating to the soccer competitions [4]. However, we stress that, contratrily to robotics, where the goal of a robot is to find an optimal solution to the problem it is facing, we try to reproduce in ARCHISIM a human behaviour that does not usually correspond to an optimal solution. An other difference is that the method chosen needs to satisfy three constraints at the same time: implement the anticipation mechanisms described by the psychologists, combine these mechanisms with a strong reactivity, and combine these two behavioural models in order to obtain credible driver agents.

6 Solving Conflicts between Driver Agents

6.1 Distributed AI Traditional Methods

Reactive Systems. The numerous reactive coordination methods used for solving conflicts, have proved themselves efficient in the domain of highly dynamical multi-agent systems. These methods are based on the use of several essential techniques [18]. Among these techniques, the most used is the one based on symmetrical force fields, which allows to define attractive and repulsive behaviours. It is applied in the area of air traffic control [30], simulation of collective animals behaviour [23, 11], or in the field of collective robotics [13].

Adding coordination actions, which allow the agents to help one another, can also ensure solving conflicts. For instance, to design autonomous aircraft controlling their traffic [10], each aircraft-agent can be aware of the team tactics and of its role in this tactics. Coordination between agents is then obtained thanks to the observation of actions of team members with respect to a set of common tactics. Another example of coordination, similarly dealing with programs of controls between reactive agents, is given by [5].

All these approaches, which are designed to be used in dynamic and unpredictable contexts, do not necessarily fit the road traffic case. In fact, their assumptions are mainly to act in reaction to the events perceived by the agent. The perception-action loop must be as short as possible, the agent can not extract too much information from its environment, nor can it anticipate in the future. And it is contradictory with the fact that driving psychologists stress the importance of anticipation even in highly perturbed contexts. The microscopic traffic simulation model, *Mitsim* [29], is a good example of the non-appropriation of a purely reactive method. This model applies a method for solving conflicts based upon the pursuit law (the vehicle acceleration is computed according to the speed of the preceding car). The decision to change lane only depends on the state of the neighbouring traffic. This limited piece of information, necessary for reactivity, induces very short-term anticipations, and produces unrealistic behaviours. Figure 3 illustrates a case of malfunction of this method.

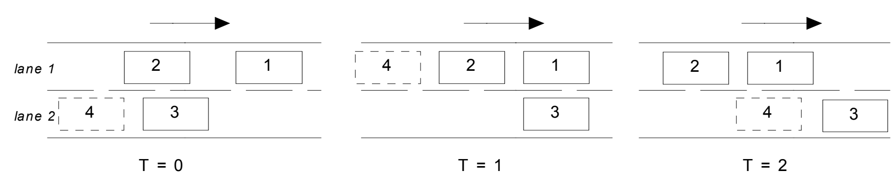


Fig. 3. The vehicles, numbered from 1 to 4, have respectively a speed of 60, 100, 80 and 90km/h. Vehicle 4 wants to reach 120km/h. At t=0, the speed of vehicle 2 is higher than that of vehicle 3. Then, vehicle 4 changes lane in order to accelerate. But, some times later, at t= 1, vehicle 2 is forced by the speed of vehicle 1. After evaluating the speed of vehicles 2 and 3 once more, vehicle 4 decides to go back to lane 2. The unuseful lane changing could be avoided if vehicle 4 anticipated in the long term, considering vehicle 1. It is what a human driver would do in this situation (according to the experiments lead by psychologists)

Furthermore, it can be difficult to escape from a conflict situation when the agents are not able to anticipate. For instance, a traffic jam around a crossroad can become so heavy that it is impossible to get rid of. This is a quite rare situation in reality and it should be so in a realistic simulation (see the example of the traffic jam formation by very simple driver-agents in [22]).

Planning Systems. Other techniques proposed in the literature, such as planning techniques, allow solving this problem by giving the agents a very high capacity to anticipate. Yet, they need some kind of communication between the agents, and also important possibilities of representation. Therefore, they are usually unable to deal with unpredictable or very complex situations in a short term (see [1]).

In this kind of systems, each agent defines a plan in which it describes the different actions it has to do at long or at short term, regarding the present status of its environment and the status it wants to reach. Given the great number of driver agents which a road traffic simulation involves, we could expect numerous plan revisions and thus a very slow individual execution time. This would prevent the agent from acting or reacting in a real time.

Moreover, if the frequency of modifications of data describing the environment is high (higher than the frequency of plan generation), the use of a planning method not only becomes inefficient, but also useless (if the agent generates plans without using them). For instance, the use of a planning technique in order to microscopically simulate a crossroad in the city of Saint-Quentin [28] allows to reproduce in a realistic way the flows which take place on the crossroads. However, the computing time becomes very important when the number of vehicles is more than one hundred.

We can conclude that, on one hand, the conflicts resolution methods which allow to react in real time in a dynamic environment, are usually not able to anticipate at a long or average term and that, on the other hand, the methods that are able to anticipate cannot cope with strongly dynamical systems. Therefore, a conflict resolution method used in the context of road traffic can be efficient only if it allows to make a good compromise between these two antagonistic criteria. It must combine reactivity and anticipation.

6.2 Proposed Method for Conflict Resolution

According to the studies led by the LPC [25], the space around the vehicle that constitutes the area of control of the driver can be split into several sectors distinguished according to their location (front, rear, left side, right side) and their proximity (very near, near, far). The driver makes his decisions on the basis of an analysis of the traffic condition in these areas. His aim is to minimize the present and future conflicts.

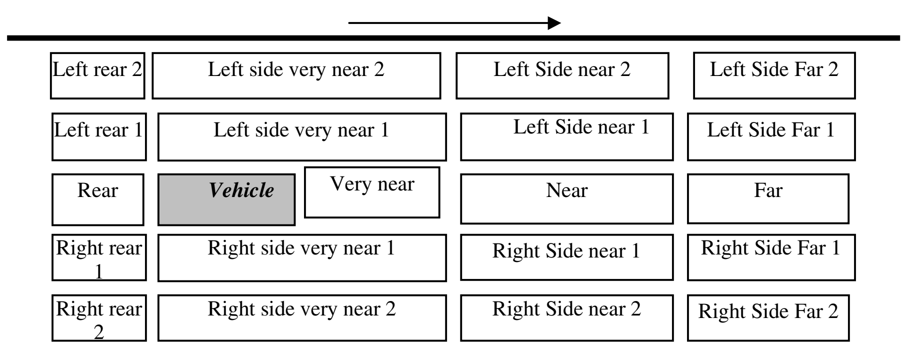


Fig. 4. The areas of perception of a simulated driver

Thus, the idea is to model the perception area of each simulated driver as a set of areas covering the field of control of a real driver (see figure 4). The decisions made by the agent driver are based on the characteristics of these areas: their infrastructure, the regulations, and the behaviour of the drivers in these areas. These characteristics are described with two parameters:

– The first is the typical interval, that measures the distribution of the speeds of the vehicles running on that area and thus the traffic stability of the area. In other words, the typical interval describes a part of the drivers' behaviour. It answers to the question whether the road users of that area have a stable behaviour or not. A high typical interval value would mean that the traffic is not stable. This would discourage the agent driver from going to that area.

– The second parameter is the area's speed. A high speed value means that the traffic on that area is fluid and encourages the agent driver to go there. This speed is computed by the agent as shown below :
  - We assume that each agent driver possesses a preferential speed which it aims to reach.
  - The speed of an area is then equal to the speed of the slower vehicle running in the area if the latter is smaller than the preferential speed of the agent driver. Otherwise its speed is equal to the agent's preferential speed.
  - However, it sometimes occurs that the geometrical dimensions of the area or the road equipements situated in the area (red light, speed limitation) impose a smaller speed than the one computed above. In that case, the area's speed will be equal to the one imposed by the road's equipements or by the area's infrastructure. For example, if an area comprises a lane closing, its speed will be equal to zero.

At each new step of the simulation, the agent driver takes information about these areas and assigns to each of them their values of stability and speed. Accordingly to these data, it decides on the actions he will undertake. The decision rules it uses are « simple ». For example in the case of file driving where the decision consists in staying or not on the same lane, the driver agent proceeds as shown below:

- Remind that, according to the psychological studies on which we are basing our work, a human driver takes the forecasting duration of the interaction into account. Likewise, an agent driver starts by assessing the duration of the interaction it is experiencing. If the duration is short it decides to adapt and to stay on its lane. For instance if the disturbing vehicle switches on its indicator, an agent driver estimates the duration of the interaction as short, since the disturbing vehicle would leave the lane soon.
- Otherwise, it associates to each of the lanes (straight-ahead, left, right) a benefit that it computes with an assessment function. This function uses the characteristics (speed and stability) of the near, very near and distant areas as parameters. The lane, which the driver will choose to drive on, is the one with the highest benefit value.

These same rules could be qualified as being reactive and anticipative at the same time. As a matter of fact, a driver agent operates in a reactive manner since its reactions only depend on its perception of the environment. The information concerning its behaviour is found in perception. To obtain these information, it does not needs to communicate with other agents nor to memorize them. Thus its reactions are fast and allow it to cope with a highly dynamical system such as road traffic.

This method also allows it to anticipate. Indeed, the agent reacts not only to the characteristics of the near areas but also on the basis of the characteristics of the distant areas. Thus it forecasts the traffic evolution by looking at distant areas and anticipating long term conflicts. Consequently, it is able to take the required decisions to avoid them. For example if the traffic is free in a near area and if an accident occurs in a distant area on its lane, the agent driver can expect the speeds to slow down and will try to avoid this interaction. In practice, the assessment function will return a low benefit value for the lane ahead and a high benefit value for the other lanes. The agent driver will choose the lane with the highest benefit value and will decide to change lane.

The distinction of the areas, according to their proximity, allows to categorize the types of actions to undertake into reactive actions and anticipating actions. This is possible because of the similarity existing in road traffic between "foreseeing" in space and "foreseeing" in time. It certainly explains that this method cannot be generalized for any multi-agent problem, but only for a class of problems still to be determined with the same space-time characteristics. The scientific interest of this method is that this idea is now accepted in many areas of human psychology. For instance, the works of psychologists on the decision making in collective sports [24], assert that once the cognitive effort of visual determination of significant visual signs has been done (which corresponds to the "experience" of the driver through a discreetization of his active perception), the implementation of simple and stereotyped rules applied to these pieces of information allows to obtain a behaviour combining anticipation and reactivity in an elegant and efficient manner. This approach is the one we already proposed in other contexts (cf. [12]), with the difference that it allows us to obtain behaviours interpreted as "credible" by the users of the simulator. The following section illustrates this point in the particular case of driving in file.

6.3 Example: Driving in File

In this example our objective was to improve the realism of the decision of lane changing, already existing in ARCHISIM, but in a poorly realistic way. It was purely reactive and, thus, showed no anticipation. To make this decision – changing or keeping lane – the agent driver must take the traffic condition not only in its near environment but also in a more distant environment into account. When the anticipations are only of short duration, which means that the traffic conditions are only taken from the near environment, the emerging behaviours are often unrealistic. We observe, in particular, a high mobility between lanes (see 3.1.1).

To validate our model at an individual level, we have applied the method of conflict anticipation to many different scenarios, which have been carefully observed by the psychologists. They have concluded that the realism of the behaviour at an individual level was greatly improved [15]. For instance, in the scenario presented at the paragraph 3.1.1, the vehicle number four does not change lane and stays on the same one.

Having a behaviour that is valid at an individual level, we wondered whether this could also lead to a valid collective behaviour. For this reason, we tested whether the traffic produced by the model complied with the general laws of traffic flow. Thus, we made a comparison between the results obtained from our simulation model and the laws described in [7]. The experiments have demonstrated that our model reproduces these laws [15].

Since, in this example, our objective was to improve the realism of the decision of lane changing, we have also validated the choice of the lane by the simulated drivers. We have based our validation on the works achieved by the Transport Engineering Department of the University of Napoli 2 (Italy) and the Faculty of Engineering of University of Reggio Calabria (Italy) [27]. On the basis of real data, the latter have described laws representing the distribution of the flow on each lane as a function of the total flow. They have also pointed out laws describing the relation between the average speed on a lane and the total flow. Both of these laws are considered in the context of a two lanes and a three lanes highway. Through the superposition of the curves obtained by simulation in ARCHISIM, and the corresponding laws [15] (see figure 5), we have been able to verify that the traffic produced by our model complied with the laws and that the assignment of flow on each lane appeared as "realistic".

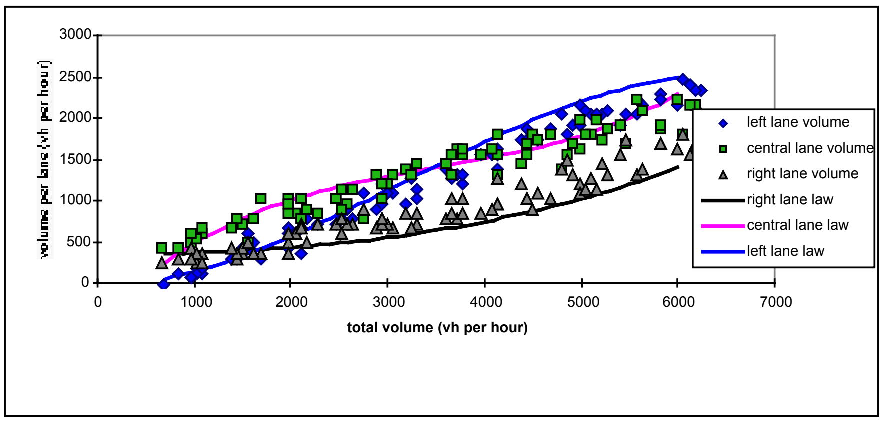


Fig. 5. Volume distribution on a three lanes highway. The laws represented on the curve were obtained through the interpolation of real data and describe the distribution of the flow according to the total flow in the case of a three lanes highway. The data represented in the form of point clouds were obtained from the application of our model to the same scenario

The lane choice analysis gives us a global validation of the changing lane decision. However, the volume distribution can be statistically correct, whereas the number of changing decisions that leads to this distribution is not. Therefore, it will be interesting to lead a more accurate validation based on the number of lane changing decisions. The only difficulty is that actual data about this criterion does not yet exist. That is why we are engaged in a cooperation with our Italian partners, during which they will eventually provide us with laws about individual lane changes.

7 Conclusion and Perspectives

The objective of ARCHISIM model is to produce realistic traffic situations by reproducing the behaviour of the drivers and putting these simulated drivers in interaction in an artificial environment. The traffic produced results from the interactions between these artificial agents (in the context described by the road infrastructure and the regulations). Yet, contrary to most of the existing models, ARCHISIM also allows to introduce in the system a real driver in a driving simulator, who receives the images of the simulated traffic. The global "realism" of a simulation (in terms of measurable data) has then to be associated with an individual "realism", (each agent being a “credible” driver for a human driver). It is under this double constraint (a constraint also strong in the design of "avatars" in virtual worlds, for example) that the agents behaviours were designed. In particular a great care was given to the mechanisms of conflicts resolution implemented by the agents, as none of the methods used in DAI was really satisfying: either because of their weak realism (reactive methods) or because they necessitated too much computation time (planning) or because they were inadequate for a very dynamic environment (planning again). The solution chosen has been inspired by the works of driving psychologists and relies on the characteristics of the road traffic, in which seeing far in space allows to make projections in time with a weak error margin. In that way we have been able to combine the design of reactive behaviours depending on a discreetization of the perception with a potentiality of anticipation, which is the implementation of the same reactive behaviour applied to distant areas. Furthermore, this solution allows us to obtain very realistic behaviour for the driving agents and has been successfully validated, with respect to a realistic traffic, in the particluar case of driving in file.

Our perspectives for the next two years are as follows:

- Refining the individual decision rules so that they will eventually be able to deal with even more complex road situations such as highway ramps, crossroads etc.
- launching a more complete set of empirical and perhaps theoretical validations (based, for example, on the number of lane changing decisions, etc.),
- generalizing this type of method to other multi-agent systems where it is crucial to combine reactivity, anticipation, real time decision-making and “credibility”. The design of artificial agents evolving in virtual reality environments together with humans or the design of robots colonies immerged in human collectivities [14] seem to be interesting problems in that respect.

References

1. Agre, P. E. and D. Chapman. Pengi: *an Implementation of a Theory of Activity*. Proceedings of AAAI'87, pp. 268-272, 1987.
2. Van Arem B., Devos A.P., Vanderschuren M.J., *The microscopic traffic simulation model MIXIC 1.3*, 1997.
3. M.Aron, approche microscopique de la cinématique des véhicules en milieu urbain, rapport INRETS n° 43, 1987.
4. Asada M., Kuniyoshi M., Drogoul A., Asama H., Mataric M., Duhaut D., Stone P., Kitano H., *«The RoboCup Physical Agent Challenge: Phase-I»,* Applied Artificial Intelligence (AAI) Journal, n° 12 (2–3), 1998.
5. Arkin, R. C. and J. D. Hobbs. *Dimensions of Communications and Social Organization in Multi-Agent Robotic Systems*. From Animals to Animats II. J. A. Meyer and S. Wilson. Cambridge, MIT Press: 486-493, 1993.
6. Buisson C., Lebaque J.-P., Lessort J.-B. STRADA, *a discretized macroscopic model of vehicular traffic flow in complex networks based on the godunov scheme*, CESE'96, Lille, France, p. 976-981, 9 au 12 juillet 1996.
7. Cohen S., *Ingénierie du trafic routier*. Cours de l'école nationale des ponts et chaussées, 1990

8. Cohen S., Semaan R., *Simulation des carrefours isolés: le logiciel CASIMIR*, rapport sur convention d'étude AFME-INRETS n° 655-8842
9. Conte R. & Castelfranchi C., *Understanding the functions of norms in social groups through simulation*. In Gilbert, G.N. & Conte, R. (eds) Artificial societies: the computer simulation of social life. London: UCL Press, 1995.
10. S.Coradeschi, L.Karlsson, *A behavior-based approach to reactivity and coordination: a preliminary report*, Intelligent agent IV Agent theories, architectures and languages 4th ATAL workshop , 1997.
11. Drogoul, A. *De la Simulation multi-agent à la résolution collective de problèmes*, Thèse de 3ème Cycle, Université Pierre et Marie Curie, 1993.
12. Drogoul, A. and D. Fresneau. Métaphore du fourragement et modèle d'exploitation collective de l'espace pour des colonies de robots autonomes mobiles. Actes des JFIADSMA'98, Paris, Hermés. 1998.
13. Drogoul, A., M. Tambe, et Fukuda, T., Eds.. *Collective Robotics*. Lecture Notes in Artificial Intelligence, Springer-Verlag, #1456, 1998.
14. Drogoul A., et Picault S. *MICRobES: vers des collectivités de robots socialement situés*, paper submited to JFIADSMA'99.
15. El Hadouaj S. ARCHISIM : Prise en compte des phénomènes d'anticipation dans le cas de conduite en file, Rapport DEA IARFA, Université Paris 6, 1998.
16. Espié S., Saad F., Schnetzler B, *ARCHISIM, Architecture parallèle multi-acteur pour simulation microscopique du trafic*, Rapport de fin de convention avec le GIE PSA-RENAULT, INRETS, 1994.
17. Espié S., *ARCHISIM, Multi_actor parallel architecture for traffic simulation*, The second world congress on Intelligent Transport Systems'95 YoKOHAMA, November 9-11, 1955.
18. Ferber J., Les systèmes multi-agents, vers une intelligence collective, InterEditions, Paris, 1995.
19. Morin J.-M. *Simaut: notice de présentation*, rapport SCE-TAUROUTE, 1989.
20. Papageorgiou M., Hadj Salem H., Elloumi N., Chrisoulakis J., Middelham F. *Integrated control of traffic corridors*, Proceedings of the 1st world congress on application of transport telematics and intelligent vehicle-highway systems, Paris, p. 556-563, 1994.
21. D. Reece and S. Shafer, *A computational model of driving for autonomous vehicles*, transpn.res-A, vol.27 A, n° 1, pp. 23-50, 1993.
22. M. Resnick, "*Turtles, Termites and Traffic Jams*", MIT Press, 1995.
23. C.W.Reynolds, Flocks, Herds and Schools: A distributed behavioral Model In computer graphics, 21(4) (SIGGRAPH'85 proceedings) pages 289-296, 1987.
24. Ripoll, H., Kerlirzin, Y., & Stein, J.F.. Analysis of information processing, decision making, and visual strategies in a complex problem solving sport situation. Human Movement Sciences., 14, 325-349. 1995.
25. F.Saad, B.Schnetzler Un modèle conceptuel du comportement des conducteurs en file sur autoroute, Rapport Interne INRETS, 1994.
26. Sichman, Conte and Gilbert (eds), *Multi-Agent Systems and Agent-Based Simulation*, LNAI series, volume 1534, Berlin, Springer-Verlag, December 1998.
27. V.Torrieri, D.Gattuso. Densità e livelli di servizio in autostrade a due e tre corsie carreggiata, 1995.
28. H. Trannois, A.Lebrun, J.L.Deleage. *Utilisation de l'intelligence artificielle distribuée pour la simulation microscopique d'un carrefour*, Recherche Transport Sécurité n° 59, Avril-Juin 1998.

29. Yang Q., Koutsopoulos, H.N. *A microscopic traffic simulator for evaluation of dynamic traffic management systems*. Transportation Research, C (43). pp. 113-119, 1996.
30. K.Zeghal. Champs de forces symétriques: Un modèle de coordination d'actions appliqué au trafic aérien, Rapport LAFORIA 93/14, Université Paris 6, 1993.

Multi Agent Based Simulation: Beyond Social Simulation

Paul Davidsson

Department of Software Engineering and Computer Science,
University of Karlskrona/Ronneby
Soft Center, 372 25 Ronneby, Sweden
Paul.Davidsson@ipd.hk-r.se

Abstract. Multi Agent Based Simulation (MABS) has been used mostly in purely social contexts. However, compared to other approaches, e.g., traditional discrete event simulation, object-oriented simulation and dynamic micro simulation, MABS has a number of interesting properties which makes it useful also for other domains. For instance, it supports structure preserving modeling of the simulated reality, simulation of pro-active behavior, parallel computations, and very dynamic simulation scenarios. It is argued that MABS is a useful technique for simulating scenarios also in more technical domains. In particular, this hold for the simulation of technical systems that are distributed and involve complex interaction between humans and machines. To illustrate the advantages of MABS, an application concerning the monitoring and control of intelligent buildings is described.

1 Introduction

Multi Agent Based Simulation (MABS) differs from other kinds of computer-based simulation in that (some of) the simulated entities are modeled and implemented in terms of agents. As MABS, and other *micro* simulation techniques, explicitly attempts to model specific behaviors of specific *individuals*, it may be contrasted to *macro* simulation techniques that are typically based on mathematical models where the characteristics of a *population* are averaged together and the model attempts to simulate changes in these averaged characteristics for the whole population. Thus, in macro simulations, the set of individuals is viewed as a structure that can be characterized by a number of variables, whereas in micro simulations the structure is viewed as emergent from the interactions between the individuals. Parunak et al. [19] recently compared these approaches and pointed out their relative strengths and weaknesses. They concluded that "…agent-based modeling is most appropriate for domains characterized by a high degree of localization and distribution and dominated by discrete decision. Equation-based modeling is most naturally applied to systems that can be

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 97-107, 2000.

modeled centrally, and in which the dynamics are dominated by physical laws rather than information processing."

We will here extend the work of Parunak et al. and argue for the applicability of MABS in other domains than it is commonly used, and while doing this compare it to some traditional simulation paradigms.

2 Multi Agent Based Simulation

MABS should not be seen as a completely new and original simulation paradigm. As we will see in this section, it is influenced by and partially builds upon some existing paradigms, such as, *parallel and distributed discrete event simulation* [16], *object oriented simulation* [23], as well as *dynamic micro simulation* [11,9].

2.1 MABS vs. Object Oriented Simulation

Since there is no commonly agreed definition of the term "agent", it is difficult to precisely define what constitutes MABS and how it should be contrasted to Object Oriented Simulation (OOS). What is referred to as an agent in the context of MABS covers a spectrum ranging from ordinary objects to full agents. For instance, we may characterize the entities in a simulation according to the following (not completely independent) dimensions:

- *pro-activeness*, ranging from purely reactive entities (cf. objects) to pro-active fully autonomous entities,
- *communication language*, ranging from having no communication at all between entities, via simple signals, e.g. procedure calls, to full agent communication languages, such as KQML [8],
- *spatial explicitness*, ranging from having no notion of space at all, to letting each entity be assigned a location in the simulated physical geometrical space,
- *mobility*, ranging from all entities being stationary to each entity being able to move around in the simulated physical space (however, not necessarily between different machines),
- *adaptivity*, ranging from completely static entities to entities that learn autonomously, and
- *modeling concepts*, ranging from using only traditional modeling concepts to using mentalistic concepts, such as beliefs, desires, and intentions.

Thus, there is no clear distinction between MABS and OOS, rather it may be viewed as a continuum; the further you go in each of these dimensions, the more MABS-like is the simulation. In an OOS, on the other hand, the simulated entities are typically purely reactive, not using any communication language, stationary, static, and not modeled using mentalistic concepts. How far you go in each of these dimensions is of course highly dependent on what entities are being simulated and the context in which they act. For instance, if a human playing soccer is being simulated, it is probably necessary to go quite far in all dimensions, whereas if a unicellular animal in a test tube is being simulated only a few dimensions are relevant.

2.2 MABS vs. Traditional Discrete Event Simulation

In principle, almost every simulation model can be seen as a specification of a system in terms of *states* and *events*. Discrete Event Simulation (DES) makes use of this fact by basing simulations on the events that take place in the simulated system and then recognize the effects that these events have on the state of the system. In continuous event simulations, state changes occur continuously in time, whereas they in DES occur instantaneously at a specific point in time. However, since it is possible to convert continuous models into discrete ones (by just considering the start and the end moments of the events), we will here only consider DES.

There are two types of DES, *time driven*, were the simulated time is advanced in constant time steps, and *event driven*, were the time is advanced based on when the next event takes place. The central structure in a traditional event driven DES is a time ordered *event list* were (time stamped) events are stored. A simulation engine drives the simulation by continuously taking the first event out of this list, setting the simulated time to the value of the time stamp of the event, and then simulate the effects on the system state (sometimes by inserting new events in the event list) caused by this event. Thus, since time segments where no event takes place are not regarded, event driven DES has the advantage of being more efficient, i.e., less time is needed to complete a simulation, than time driven DES. On the other hand, since time is incremented at a constant pace, e.g., in real time, during a simulation in time driven DES, this is typically a better option if the simulation involves human interaction (or even just monitoring) at run time, e.g., in training situations.

If we compare MABS to traditional DES we find that it has several advantages. Just like OOS, it supports structure preserving modeling and implementation of the simulated reality. That is, there is a close match between the entities of the reality, the entities of the model, and the entities of the simulation software. This simplifies both the design and the implementation of the software, and typically results in well-structured software. In addition, we argue that MABS has the following important advantages compared to more traditional DES techniques:

- It supports modeling and implementation of pro-active behavior, which is important when simulating humans (and animals) who are able to take initiatives and act without external stimuli. In short, it is often more natural to model and implement humans as agents than objects.
- It supports distributed computation in a very natural way. Since each agent is typically implemented as a separate piece of software corresponding to a process (or a thread), it is straight-forward to let different agents run on different machines. This allows for better performance and scalability.
- Since each agent typically is implemented as a separate process and is able to communicate with any other agent using a common language, it is possible to add or remove agents during a simulation without interruption. And, as a consequence of this and the structure preserving mapping between the simulation

software and the reality, it is even possible to swap an agent for the corresponding simulated entity, e.g., a real person during a simulation. This enables extremely dynamical simulation scenarios.
- It is possible to program (or at least specify) the simulation model and software on a very high level, e.g., in terms of beliefs, intentions, etc., making it easier for non-programmers to understand and even participate in the software development process.

Of course, there are also some disadvantages with MABS compared to DES. For instance, a fully agent-based approach typically uses more resources, both for computation and communication, which may lead to less efficient (slower) simulations. Also, whereas MABS is very appropriate for time driven simulations, it is less appropriate for event driven simulations. In event driven MABS there is a need for either a central coordinator that keeps track of which event to be executed next, or a large amount of synchronization between the agents. Having a central coordinator would be contrary to some of the ideas that motivated a multi agent based approach in the first place, and the synchronization would slow down the simulations considerably.

2.3 MABS vs. Dynamic Micro Simulation

The purpose of Dynamic Micro Simulation (DMS) is to simulate the effect of the passing of time on individuals. Data from a large random sample from some population is used to initially characterize the simulated individuals. Some possible sampled features are, e.g., age, sex, and employment status. A set of *transition probabilities* is used to simulate how these features will change over a time period. The transition probabilities are applied to the population for each individual in turn, and then repeatedly re-applied for a number of simulated time periods.

Compared to MABS, DMS has two main limitations. First, the behavior of each individual is modeled in terms of probabilities and no attempt is made to justify these in terms of individual preferences, decisions, plans, etc. Second, each simulated person is considered individually without regard to its interaction with others. Better results may be gained if also cognitive processes and communication between individuals were simulated and by using agents to simulate the individuals, these aspects are supported in a very natural way.

In the past, both MABS and DMS has been applied mostly in purely social contexts [10], e.g., to validate or illustrate social theories (including biological, economic, and political theories), or predict the behavior of interacting social entities. Examples of such domains are:

- actors in financial markets [1]
- consumer behavior [13]
- people in crowds [21] and animals in flocks [20]
- animals and/or plants in eco-systems [5, 6]
- vehicles (and pedestrians) in traffic situations [22]

In most, if not all, of these simulation scenarios, only social entities are present. The main advantage of MABS explored in these simulations is that it facilitates the simulation of group behavior in highly dynamic situations, thereby allowing the study of "emergent behavior" that is hard to grasp with macro simulation methods. MABS has proven to be well suited for the simulation of situations where there are a large number of heterogeneous individuals who may behave somewhat differently and is therefore an ideal simulation method for the social sciences.

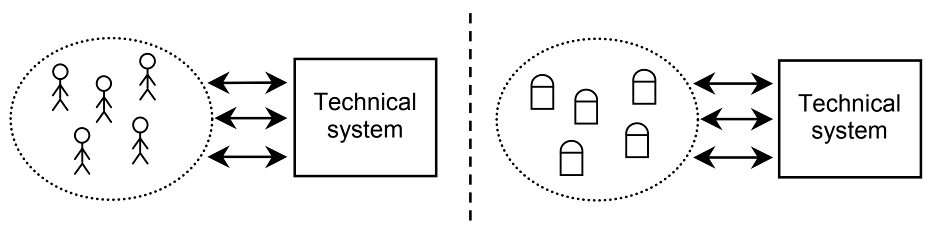


Fig. 1. To the left: the fielded system used by people. To the right: agent-based simulation of people using the system

However, as we have seen there are a number of further advantages of MABS compared to traditional simulation techniques. This suggests that MABS may be a useful technique also for other types of simulation than of purely social systems. We argue that MABS is particularly useful for simulating scenarios in which humans interact with a technical system. (A similar argument has been made by Moss et al. [17] in the context of simulating climate change were humans interact with a physical system.) The purpose of such simulations could then be, e.g., evaluation of the technical system, or for training future users of the system. As a case study, an evaluation of a "socio-technical" system concerning the controlling of intelligent buildings will be described.

Many new technical systems are distributed and involve complex interaction between humans and machines. The properties of MABS discussed above makes this technique especially suitable for simulating this kind of systems. As illustrated in Fig. 1, the idea is to model the behavior of the human users in terms of software agents. In particular, MABS seems very suitable in situations where it is too expensive, difficult, inconvenient, tiresome, or even impossible for real human users to test out a new technical system.

Of course, also the technical system, or parts thereof, may be simulated. For instance, if the technical system includes hardware that is expensive and/or special purpose, it is natural to simulate also this part of the system when testing out the control software. In the next chapter we will see an example of such a case, a simulation of an "intelligent building".

3 A Case Study (Evaluation)

In a de-regulated market the distribution utilities will compete with added value for the customer in addition to the delivery of energy. We will here describe a system consisting of a Multi-Agent System (MAS) that monitors and controls an office building in order to provide services of this kind. The system uses the existing power lines for communication between the agents and the electrical devices of the building, i.e., sensors and actuators for lights, heating, ventilation, etc. The objectives are both energy saving, and increasing customer satisfaction through value added services. Energy saving is realized, e.g., by lights being automatically switched off, and room temperature being lowered in empty rooms. Increased customer satisfaction is realized, e.g., by adapting temperature and light intensity according to each person's personal preferences. A goal is to make the system transparent to the people in the building in the sense that they do not have to interact with the system in any laborious manner. By using an active badge system [12], the MAS automatically detects in which room each person is at any moment and adapts the conditions in the room according to that person's preferences. This project is currently in its simulation phase, but some fielded experiments at our test site, the Villa Wega building, in Ronneby Sweden, have been made to assure that the performance of power line communication is sufficient for controlling, e.g., radiators.

3.1 The Multi-Agent System

Each agent corresponds to a particular entity of the building, e.g., office, meeting room, corridor, person, or hardware device. The behavior of each agent is determined by a number of rules that express the desired control policies of the building conditions. The occurrence of certain events inside the building (e.g., a person moving from one room to another) will generate messages to some of the agents that will trigger some appropriate rule(s). The agents execute the rule(s), with the purpose to adjust the environmental conditions to some preferred set of values. The rule will cause a sequence of actions to be executed, which will involve communication between the agents of the system. For the format of the messages a KQML-like [8] approach was adopted. The language used to implement the MAS is April [15]. The agent-based approach provides an open architecture, i.e., agents can be easily configured and even dynamically re-configured. It is possible to add new agents or change their behavior at run-time without the need of interrupting the normal operation of the system.

There are four main categories of agents in the MAS: *Personal comfort agents*, which corresponds to a particular person. It contains personal preferences and acts on that person's behalf in the MAS trying to maximize the comfort of that person. *Room agents*, which corresponds to and controls a particular room with the goal of saving as much energy as possible. *Environmental parameter agents*, which monitors and controls a particular environmental parameter, e.g., temperature or light, in a particular room. They have access to sensor and actuator devices for reading and changing the parameter. Finally, the *badge system agent* keeps track of where in the building each person (i.e., badge) is situated. More details about the MAS can be found in [2].

Typically, the goals of the room agents and the personal comfort agents are conflicting: the room agents maximizing energy saving and the personal comfort agents maximizing customer value. Another type of a conflicting goal situation would be the adjustment of temperature in a meeting room in which people with different preferences regarding temperature will meet. We experimented with different approaches to conflict resolution, the simplest being based on a priori reasoning. For instance, the Room agents determine the desired temperature in a room by just accepting the temperature preferred by the person in the room. If many persons are in the room, it either takes the average of the preferred values, or makes use of priorities, e.g., by taking into account only the preferences of the manager and/or visitors. Of the run time solutions to conflict resolution, "coin flipping" using a random number generator is the simplest. A more sophisticated approach is to make use of a mediator, i.e., a third agent able to make an objective assessment of the situation, to resolve the conflict. We made some initial experiments using *pronouncers* [3] as mediators. Finally, we also regarded the possibility of resolving conflicts using negotiation between the agents. For example, an agent may propose that "if this time my preferences are used, yours will be used next time we are in the same room."

3.2 Evaluation of the MAS

Since it would be quite expensive to equip the Villa Wega building with all the necessary hardware in order to evaluate the approach outlined above, we decided to make a preliminary evaluation of the approach (i.e., the MAS) through simulations. In this case, the technical system can be divided into two parts; the hardware, i.e., the building including sensors and effectors, and the software, i.e., the MAS. Thus, we simulate the hardware and let the actual MAS, which will be used in the fielded application, interact with it instead of the actual hardware. Please note that the MAS does not simulate anything, it "just" monitors and controls the building.

Now, we must also simulate the people working in the building. As indicated earlier, we may do this by MABS where each person corresponds to an agent. This agent simulates the behavior of that person (to be contrasted to the personal comfort agents in the MAS which serves the person, i.e., is an agent in the true sense of the word). Fig. 2 illustrates the different parts of the simulation software.

By just specifying a few parameters that characterize the behavior of a person (e.g., which rooms she/he normally visits and how often, and the mean value and standard deviation for the time when certain events takes place, e.g., arrival to the building), we can easily create an arbitrary number widely different simulated persons. As we also want to simulate the building without the MAS in order to estimate the amount of energy saving and increased personal comfort the MAS can achieve, some additional parameters are needed, e.g., the person's tendency to forget to turn off the lights etc. A presentation of the simulation scenarios and the results can be found in [4].

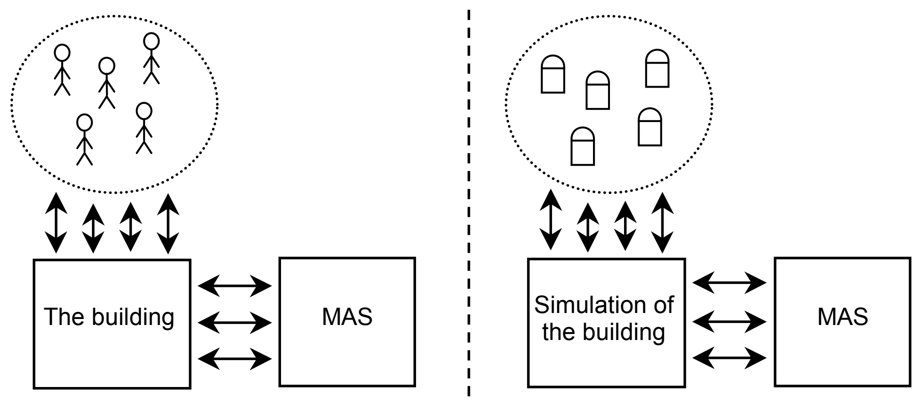


Fig. 2. Fielded (left) and simulated (right) use of the intelligent building control system

The simulation of the physical properties of the building was based on the thermodynamical models described by Incropera and Witt [14], which were discretized according to standard procedures (cf. Ogata [18]). All the thermodynamical characteristics of a room are described by two constants: the thermal resistance, *R*, which captures the heat losses to the environment, and the thermal capacitance, *C*, which captures the inertia when heating up/cooling down the entities in the room. (In all simulations below we use the sample time 1 minute.). The temperature, T_{xi}, in room *x* at time *i* is described by:

$$T_{xi} = \frac{1}{1+\frac{1}{R_x C_x}}\left(T_{x(i-1)} + \frac{P_i + \frac{T_{outi}}{R_x}}{C_x} \right) \tag{1}$$

where P_i is the heating power, T_{outi} the outdoor temperature, and $T_{x(i-1)}$ is the temperature one minute ago. Thus, the dynamics of each room is simulated using a traditional equation-based model, indicating the possibility integrating different simulation paradigms in order to explore their respective strengths.

4 Concluding Remarks

In the last chapter we gave a high-level description of a project aimed at investigating the usefulness of multi-agent systems for the design of control systems for intelligent buildings. The purpose of this case study was to argue for the use of MABS when *evaluating* complex technical system that are distributed and involves interaction with humans. A number of advantages of MABS can be identified, e.g.:

- Since each person is simulated by a separate agent, it is easy to simulate persons with very different behavioral characteristics.

- It is not necessary create a long event list prior to the simulation. The pro-active behavior of people moving from one room to another etc. is easily achieved. Only some parameters describing the simulated person's behavioral characteristics is needed.
- Well-structured simulation software.
- It is easy to increase performance since different groups of people may be simulated on different machines (also supports scaling).
- Very flexible simulation scenarios can be constructed since it is easy to add another person to (or remove one from) the scenario during a simulation.

In the case study, the evaluation of customer satisfaction was rather primitive. Although MABS probably is the most suitable simulation technique for making this kind of evaluation, it is difficult to define a truly meaningful metric for customer satisfaction. The best we can do is to continually measure the difference between the desired values of the relevant environmental parameters (according to the preferences specified by the person in question) and the actual values of those parameters during a simulation. However, we believe that there are more subtle aspects that influence the satisfaction a person gets from a system such as this. Unfortunately, these are probably difficult to define explicitly (and therefore hard to measure) but are at least as important. One such aspect regards personal integrity. How comfortable is it to know that your manager may know exactly where you are at any time? Thus, it seems difficult to make such an evaluation based only on computer simulations; it is necessary to let real persons use the system. Note, however, that this is not a limitation only for MABS, but for computer simulations in general.

We have not here demonstrated the usefulness of MABS for the purpose of *training* people. However, it is not difficult to find domains in which MABS seem to have a great potential, e.g., car driving [7], managing troops and other military units, managing companies, etc.

Acknowledgement

The author wishes to acknowledge the valuable contribution from the colleagues in the ALFEBIITE project.

References

1. Basu N. and Pryor R.J.: Growing a Market Economy. Technical Report SAND-97-2093, Sandia National Laboratories, Albuquerque, NM, USA (1997)
2. Boman M., Davidsson P., Skarmeas N., Clark K., and Gustavsson R.: Energy Saving and Added Customer Value in Intelligent Buildings. In: *Third International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology* (1998) 505-517

3. Boman M., Davidsson P., and Younes H.L.: Artificial Decision Making under Uncertainty in Intelligent Buildings, *Fifteenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, (1999) 65-70
4. Davidsson P. and Boman M.: Saving Energy and Providing Value Added Services in Intelligent Buildings: A MAS approach. In: *Agent Systems, Mobile Agents, and Applications*, Springer Verlag (2000) 166-177
5. Deutschman, D.H., Levin S.A., Devine C., and Buttel L.A.: Scaling from Trees to Forests: Analysis of a Complex Simulation Model. In: *Science Online: An electronic journal* (1997) (www.sciencemag.org)
6. Drogoul A., Corbara B., and Fresneau D.: MANTA: New experimental results on the emergence of (artificial) ant societies. In: Gilbert N. and Conte R. (eds.), *Artificial Societies: the computer simulation of social life*, UCL Press (1995)
7. El hadouaj S., Drogoul A., and Espié S.: To combine reactivity and anticipation: the case of conflicts resolution in a simulated road traffic. This volume.
8. Finin T., Fritzson R., and McKay D., et al.: An Overview of KQML: A Knowledge Query and Manipulation Language. Technical report, Department of Computer Science, University of Maryland, Baltimore County, USA (1992)
9. Gilbert N.: Computer Simulation of Social Processes. *Social Research Update*, Issue 6, Department of Sociology, University of Surrey, UK (1994) (www.soc.surrey.ac.uk/sru/)
10. Gilbert N. and Conte R. (eds.): *Artificial Societies: the computer simulation of social life*. UCL Press (1995)
11. Harding A. (ed.): *Microsimulation and Public Policy*. Elsevier (1996)
12. Harter A. and Hopper A.: A Distributed Location System for the Active Office. In: *IEEE Network* 8(1) (1994)
13. Hämäläinen R.P.: Agent-Based Modeling of the Electricity Distribution System, In: *IASTED International Conference on Modelling, Identification and Control*, (1996) 344-346
14. Incropera F.P. and Witt D.P.: *Fundamentals of Heat and Mass Transfer (3rd edition)*. Wiley and Sons (1990)
15. McCabe F.G. and Clark K.L.: April: Agent Process Interaction Language. In Wooldridge M. J. and Jennings N. R. (eds.), *Intelligent Agents*, Springer Verlag (1995) 324-340
16. Misra J.: Distributed Discrete-Event Simulation. *ACM Computing Surveys*, 18(1) (1986) 39-65
17. Moss S., Pahl-Wostl C., and T. Downing: Agent Based Integrated Assessment Modelling: The example of Climate Change, *Integrated Assessment*, 1, 2000.
18. Ogata K.: *Modern Control Engineering (2nd edition)*. Prentice-Hall (1990)
19. Parunak H.V.D., Savit R., Riolo R.L.: Agent-Based Modeling vs. Equation-Based Modeling: A Case Study and Users' Guide. In Sichman, J.S., Conte, R., Gilbert, N. (eds.), *Multi-Agent Systems and Agent-Based Simulation*, Springer Verlag (1998) 10-26
20. Reynolds C.W.: Flocks, Herds, and Schools: A Distributed Behavioral Model", Computer Graphics, 21(4) (1987) 25-34

21. Williams R.J., “Simulation for public order training and preplanning. In Kanecki D. (ed.), *Emergency Management Conference* (1994) 61-66
22. Yang, Q.: A Simulation Laboratory for Evaluation of Dynamic Traffic Management Systems. Ph.D. Thesis, Center for Transportation Studies, Massachusetts Institute of Technology, USA (1997)
23. Zeigler B.P.: *Object Oriented Simulation with Hierarchical Modular Models*. Academic Press (1990)

A Multi-Agent Based Simulation of Sand Piles in a Static Equilibrium

Laurent Breton[1], Jean-Daniel Zucker[1], and Eric Clément[2]

[1] LIP6 - UPMC/CNRS - 4,
Pl. Jussieu - Case 169 - 75232 Paris Cedex 05
{Laurent.Breton,Jean-Daniel.Zucker}@lip6.fr
http://www-poleia.lip6.fr/~{breton,zucker}
[2] Laboratoire des Milieux Désordonnés et Hétérogènes - Université Paris VI
4, place Jussieu - case 86 - 75005 Paris
erc@ccr.jussieu.fr

Abstract. In granular physics, numerical simulations are often used to better explain the link between microscopic interaction mechanisms (e.g. contact forces between beads) and macroscopic laws (e.g. the maximum load at the bottom of a silo). Here we study the mechanical properties of a pile such that all the grains are in a static equilibrium. This paper presents a multi-agent approach that solves this problem. The main idea of this algorithm is to consider a grain as a reactive agent which goal is to reach its own static equilibrium. This algorithm can compute piles in practice in a complexity that is linear with respect to the number of grains. The performance of our algorithm has paved the way for the development of GranuLab, a virtual laboratory for scientific experiments in the domain of granular physics. We present both the algorithm and our validation processes.

1 Introduction

Today, about 80% of the final (or intermediate) products of process industries are presented as grains or powders. However, the connection between grain interaction mechanisms and reliable macroscopic laws governing their mechanical aspects has not been established. Recent research shows that the heart of the problem is located in the imperfect or even non-existent understanding of relations between local intergrain interactions (contact forces) and the possibility to form self-organized structures at an intermediate level (force chains, arches). These structures have a dynamics that is governed by the action of external forces, by the dynamics of local contact laws and by the variation of piling structures. Recent measurements ([9], [10], [7], [19], [20]) and numerical simulations by discrete elements ([12], [4], [11]) have shown that these structures describe the piling history and define some macroscopic parameters such as force fields and distortions. Their dynamic is also at the origin of the constraint fluctuations observed in practice ([9], [10], [7], [19], [20]).

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 108-118, 2000.

Physicists and artificial intelligence researchers have been working together to address the above problems. The main objective of this collaboration is to design and develop a virtual laboratory for scientific experiments ([1], [14], [18]) in granular physics : GranuLab. This laboratory must help the researcher to understand the different aspects of force network dynamics. The first and necessary stage in the design of this environment is the creation of a problem-solving method which must be efficient for reasonable size experimentation and reliable in order to be really used as a basis for research in physics.

Part two presents the problem as physicists see it. We show the numerical approaches they proposed and explain why it is necessary to resort to such approaches. Then, we present the multi-agent model and the eco-solving algorithm we have designed and implemented. Finally, the way to validate results obtained using the multi-agent approach will be explained. This validation is essential, as it precedes the use of this environment as a virtual experimentation laboratory .

2 Physical Problem of Granular Piling

2.1 Reasons for the Research

Contrary to the mechanical study of a solid, granulars have the nasty habit of behaving in unexpected ways. Let's take the example of grain silos: the out-flow is often stopped by the creation of granular arches (cf. Fig. 1). This arch is very strong and bears a large part of the weight in the silo, and it is difficult to break it. So far, engineers haven't found any solution other than to hammer the side of the silo.

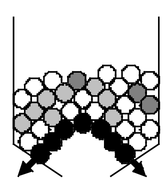


Fig. 1. Arch in a silo

To make the problem more difficult, the grain behaves like a liquid and suddenly flows downwards when the arch is broken. With the creation of such arches also comes the question of the maximal load supported by the side of the silo where the arch is located. Engineers who design such silos base their work on existing theories of continuum environments and on their experimental knowledge in order to evaluate the required strength of the materials and to provide comfortable safety margins. Understanding these flow phenomena and being able to predict such disasters and their consequences are therefore vital. A lot of research in physics aims to better understand the force networks underlying granular equilibrium.

2.2 Real Experiments

The first experiments to show the presence of a force network with this kind of pile were based on the theory of photoelasticity [15]. The piling of plexiglass cylinders was lit up using crossed polarizers. With no action, the pile seems to be dark, but under the action of strong forces, some grains bend and appear locally more luminous (cf. Fig. 2.a). However, the distortion of the cylinder must be sufficient to let the light through.

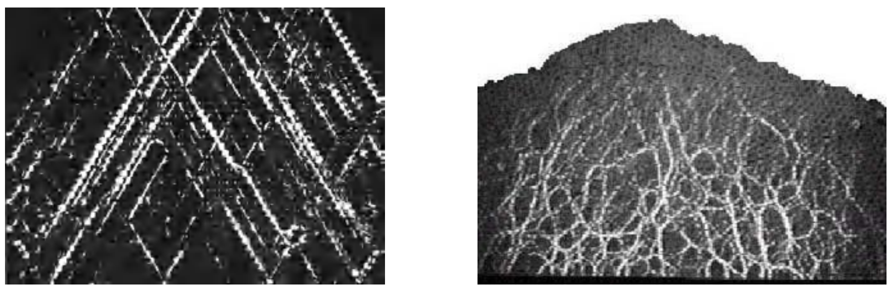


Fig. 2. Force networks **a)** piling under pressure and **b)** piling supporting its own weight

It is only recently that Behringer's team at Duke University visualized force networks in a pile supporting its own weight, using very sensitive materials (cf. Fig. 2.b). Generally, the macroscopic measurements of forces taken on real pilings require drastic experimental conditions, in order to obtain values that can be reproduced [19]. Large fluctuations are observed (about the mean [8]) because of the finite size of force sensors ([9], [10], [7]). These measurements are not very helpful concerning the internal structure of force networks, and in order to understand these macroscopic phenomena numerical simulation is essential [12]. Therefore, these experiments have a key-role in the validation of these models designes by physicists.

2.3 A Simplified Physical Model

Currently, many physicists specializing in granular environments are interested in discovering an efficient physical model to explain the above-mentioned phenomena. However, this research comes up against physical and numerical problems. First of all, which scale should be used: microscopic, at the grain level or at the force-line? Another difficulty is the nature of the physical mechanisms involved. Do we need to model electrostatic interactions, vectorial forces, or is there a dynamics of inter-grain points of contact ? From a numerical point of view, the number of grains necessary for a statistical study of the model, the hysteretical nature of the piling and the heterogeneity of materials make simulations very difficult. These difficulties have led us to consider a pile model that can give account of certain macroscopic phenomena, the solutions of which can be computed.

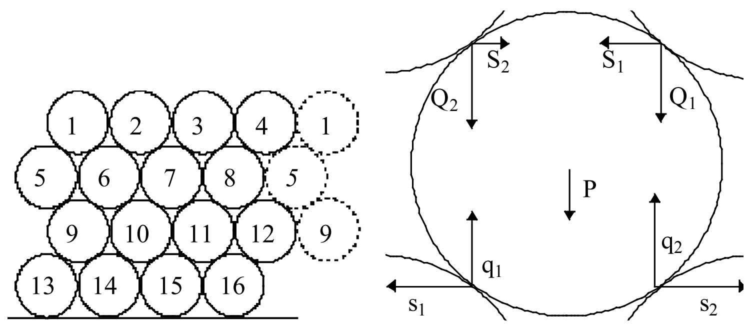


Fig. 3. a) Cannonball pile, **b)** forces applied to the grain

Clement suggests studying bidimensional piles in hard, regular spherical cannonballs (cf. Fig. 3.a) [4]. The pile is torodal (periodic limit conditions) in order to be free from limit conditions due to the sides of the silo (cf Fig. 3.a, grain number 5 will have as neighbors grains number 1, 4, 12 and 9). In addition, lateral contacts are deliberately ignored (cf. Fig. 3.a, there is no contact between grains 5 and 6).

$$\begin{cases} Q_1 + Q_2 + q_1 + q_2 = -P & (1) \\ S_1 + S_2 + s_1 + s_2 = 0 & (2) \\ Q_2 - Q_1 + \tan\theta\,(S_1 + S_2) = q_2 - q_1 + \tan\theta\,(s_1 + s_2) & (3) \end{cases}$$

Fig. 4. Equilibrium by **1)** vertical translation, **2)** horizontal translation, **3)** rotation

In order for the pile to be in static equilibrium, each grain must counter-balance the vectorial forces that are applied to it. Every grain has to satisfy the same system of equations (cf. Fig. 4). The Q (resp. S) represent the vertical projection (resp. horizontal) of the contact force for q = 60□ In addition, a constraint is added at the level of contact forces so that they are contained in their Coulomb cone of friction which defines the static limit of contact for solid upon solid. Each grain must therefore have a set of forces that satisfies both its equilibrium and the friction constraints on its four contact points. But, since the condition of Coulomb, which is valid in every contact, is a double inequality, a pile will have a large number (infinite in fact) of solutions that satisfy all conditions of static equilibrium.

2.4 Stochastic Approach to Equation Solving

The numerical method MC-Granu used by Eloy et al. [4] for the complete solution of equations of the static equilibrium of a pile is a stochastic one. It is similar to Monte-Carlo methods which make random choices of variables. The problem-solving process is top down through random sampling of layers. First of all, we are going to show how the equilibrium of grains is reached and then we will explain the algorithm globally. When solving from top to bottom, we consider that the higher values of grain forces (cf. Fig. 4, Q1, S1, Q2 and S2) are fixed, so we have a system with 3 equations and 4 unknowns (q1, s1, q2 and s2). The propagation of these constraints from the limit conditions at the top of the pile brings only very little information to the lower layers and does not reduce the size of the search space [6].

In MC-Granu, we have to choose a parameter p for each grain; this parameter defines how the load is to be distributed to the lower grains (cf Fig. 5).

```
While the last layer is not reached, Do
    1. Create a sample of the current layer, of size N,
       while choosing randomly the parameter p for all
       the grains of the layer.
    2. Forward checking: eliminate from the sample the
       layers that generate failures to the following
       layer.
    3. If the sample is empty,
       Then backtrack 2^(failures-1) layers and increment the
       number of failures
       Else choose randomly a layer from the sample and
       propagate its values to the following layer.
    4. If we arrive at a layer that has not yet been
       reached
       Then discount to zero of the number of failures.
End while
```

Fig. 5. MC-Granu algorithm

Unfortunately, it often happens that all layers generate failures to the following layers among those sampled. In this case, MC-Granu proceeds to a backtrack proportional to the number of failures encountered ($2^{failures - 1}$). Thus, the pile can be entirely rebuilt hundreds of times, before having the chance of generating a pile with few failures. In practice, this algorithm has exponential complexity with the width of the pile (cf. §4.2). The wider the pile, the more difficult it is to generate randomly layers without any failures.

2.5 The Problem to Be Addressed

Using MC-Granu algorithm, it isn't possible to generate piles more than fifty grains in width, which makes statistical studies of the model difficult. What we need is a problem-solving algorithm that can be used by researchers in granular physics and that is generic and valid. The algorithm must be fast in order to be able to generate, in a reasonable time, piles reaching several hundred grains in width; this is the necessary condition for all statistical studies of the model. Moreover, a generic algorithm would allow researchers to modify or to complicate their model without having to redesign the system entirely. Finally, to be validated, the algorithm would have to corroborate the macroscopic physical phenomena we are trying to understand.

In granular physics, researchers are interested mainly in the macroscopic behaviors of the force network. As the fluctuations of such parameters are such that the value of their order of magnitude is the mean, it is necessary to generate tens of thousands of solutions to be able to do a real statistical study. In addition, the generated piles must be deep enough to be free from first layer fluctuations.

3 Distributed Modeling

It was therefore necessary to find a usable problem-solving method. To do so, we suggest a multi-agent approach for the numerical simulation of the force networks in a pile of sand.

The idea of using multi-agents systems (MAS) for simulation in physics is very recent. In a theoretical study of the contributions of MAS for physics, Treuil et al. [16] introduced the notion of “simulons” to describe agents simulating physical processes. Simulons are not the physical processes themselves, but expert agents who reason on these physical processes. They identify three characteristics of a simulon agent: the possibility to distribute the calculations, the autonomy of the agent and its capacity to communicate. Servat adopted this distinction in his study of flowing water droplets by a multi-agent system [13]. Each ball of water being autonomous, it computes its trajectory itself and these calculations are distributed between agents. In the same way, their capacity to communicate allows them to regroup in pools according to the communicating vessels principle.

For the problem we are interested in here, considering a grain as an agent allows the problem-solving of the pile to be decentralized locally to each grain. This is because the search space is too wide to be solved globally, whereas locally possibilities offered to each grain are given by its equations. Besides, our agents are sufficiently autonomous to be able to introduce dynamics in the topology (avalanches, silo out-flows, etc.) or other modifications to the model. Finally, their capacity to communicate allowed us to define simple problem-solving behaviors, in collaboration with the physicists.

3.1 Model of the Grain-Agent

Because of the inertness of a real grain, the modeling of the grain by an essentially reactive agent is natural. In order to design the grain-agent, we used the Cassiopée MAS design methodology [1], which defines an agent in five incremental layers, starting from its different roles. The interest of an incremental construction is twofold. First, it is possible to modify an external layer without changing anything of the internal layers; second, this approach respects the principle of parsimony: we can stop to complicating the definition of the agent as soon as the system computes the desired function.

The first level defines the agent's individual role. In our case, it is the role of a single grain, that does not yet have any neighbors but is already a physical entity, a grain in the pile environment. Its knowledge is limited to its geometry and its set of forces. Its role is first to initialize randomly its set of forces and then to solve itself: considering a set of fixed forces, it computes a possible value for the other forces.

```
Solve_Individual(x, set of fixed forces)
1.Compute_Intervals(x, non fixed forces)
2.If No_Solution(x), return FALSE
  Else    If Is_Solved(x) return TRUE
          Else Choose_Interval_Values(x, non fixed
            forces)
3.Return TRUE
```

Fig. 6. Individual solution of the grain

The function Compute_Intervals(x, set of forces) computes the possible values for the forces, according to the equations of the model. The function No_Solution(x) returns TRUE if the equilibrium is impossible. Then we choose a random value for each non fixed force using the function Choose_Interval_Values(x, non fixed forces). We say that a grain may be solved independently of its neighbors. Here, the grain tries to satisfy its individual equilibrium.

The second level defines the agent's topology. It is an intermediate level between its individual role and its social role. Here the grain determines its neighbors, those with which it will be able to communicate force values afterwards. Currently, the topology of the grain is static, at four points of contact, but we will eventually need to consider a more complex model.

The third level defines the agent's social role[1], its behaviors with respect to its neighbors (cf. Fig. 7). The grain will try to solve itself socially, by accepting forces that its neighbors send it and repelling others in order to be able to remain locally in static equilibrium.

```
Solve_Social(x)
1.If Solve_Social(x), return TRUE
2.Else,   Forces_wanted = Accept_Force(x)
          Solve_Individual(x,Forces_wanted)
          For all modified forces, Do
             Repel(y,F)
          EndFor
3.        return TRUE
```

Fig. 7. Social solution of the grain

The function Accept_Force(x) accepts the forces of its neighbors and sends back the list of fixed forces. The function Repel(y, F) sends the force F to its neighbor and warns it that it must solve itself socially.

The fourth level defines possible grouping of agents. Though it has not been necessary in this model, we plan to extend the model to capture grain layers, lines of force, arches and other intermediate entities of interest to physicists.

The fifth, collective, level defines the behavior of theses groups among themselves. This methodology is recursive which reconcile the top-down and bottom-up approaches of MAS. In one hand, the top-down approach allows the designer to build groups and abstractions. This construction can be made recursively and more and more accurately till indivisible entities. In the other hand, the bottom-up approach supposes the designer to build more and more complex entities till the desired upper level of abstraction. The Cassiopea methodology allows the designer to create groups and abstractions but always under constrainst of the lower level of the agent.

3.2 The Eco-Solving Algorithm

The local solving process of each grain having been defined, we are now going to show the global problem-solving of the pile (cf. Fig. 8.). The pile initializes grains with a random set of forces and memorizes a list of grains not yet socially solved.

[1] The term "social" follow the Cassiopée methodology for generic agents and doesn't have the same meaning as in Social Sciences.

First, they are all solved individually but not socially because each one is in equilibrium with its own set of forces, without regard to the contact forces its neighbors have chosen, which means that we violate Newton's law of action-reaction.

```
Eco_Solving(pile, Heuristic_Choice())
  While Not_Solved(pile) Do
    grain = Heuristic_Choice(pile)
    Solve_Social(grain)
    update_List(pile,grain)
  End While
```

Fig. 8. Eco-solving Algorithm

The function Not_Solved(pile) checks whether the list of grains not yet solved is empty. The function Update_List(pile, grain) updates the list of grains not yet solved, while looking at its neighborhood. The way the pile is solved only depends on a heuristic, given as a parameter of the algorithm.

4 Experiments and Results

4.1 Methods

One of the main goals of the experiments is to determine the most suitable eco-solving heuristic for the problem. We have therefore defined several heuristics to choose the grain that needs to get priority treatment, each one corresponding to a distinct physical notion: the choice of the grain bearing the maximum pressure (resp. minimum) from its neighbors, random choice of grain, the first grain encountered when scanning the pile from top to bottom or the grain with the least (resp. the most) freedom to move all constitute possible heuristics. All these heuristics have eventually succeeded in solving piles.

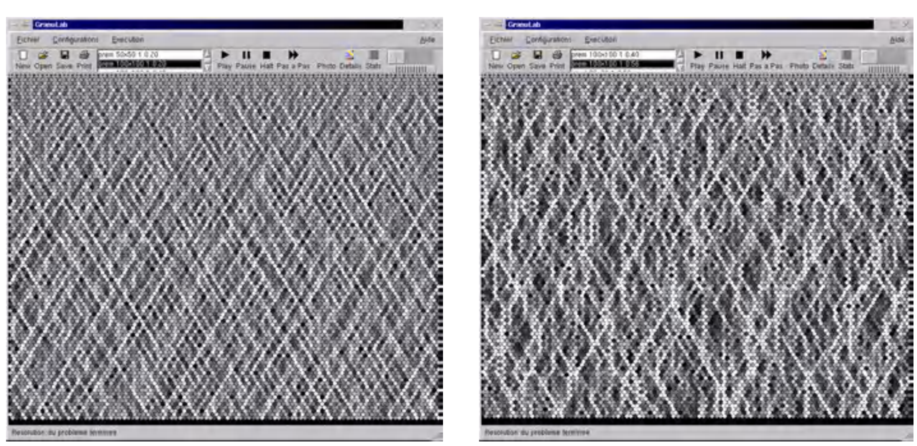


Fig. 9. Disorder of the force network, pile 100x100 and mu = **a)** 0.2 and **b)** 0.6

But in order to validate our eco-solving algorithm, we must also check the presence of certain macroscopic behaviors known from real piles. For example, we have shown that the higher the friction coefficient, the more disordered the force network. We have also found the typical patterns of a force network in simulated piles: lines of forces and arches (cf. Fig. 2 and 9).

4.2 Results

In its present version GranuLab (cf. Fig. 9a) can generate a problem pattern associated with a problem-solving method. The problem configuration memorizes the geometry of the grains, the number and size of the piles to be computed, the heuristic to be used for the eco-solving, etc. Then, we can create a new instance of this problem and launch the process, for example, step by step and with display in order to observe the problem-solving behavior of the heuristic chosen.

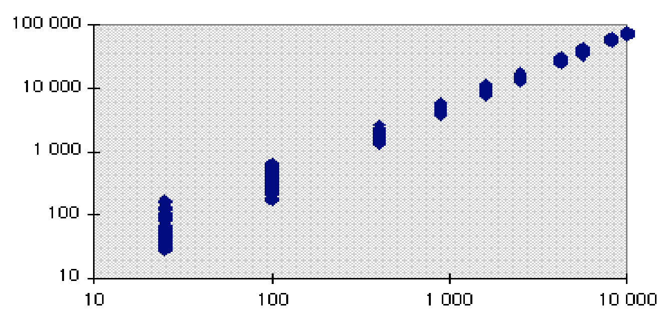


Fig. 10. Number of iterations by number of grains in a log-log scale

All the heuristics listed above have succeeded in solving piles, but "eco-serpentine", the one that chooses the first unsolved grain encountered while scanning top-down, gave the best results. In practice, the complexity of eco-serpentine is linear with respect to the number of grains. Thus, we have solved piles up to a million of grains (1000x1000) in reasonable computing time (cf. Fig. 10 and 11).

Table 1. Table showing the average performance of algorithms for 100 solved piles

Heuristic	Eco-Serpentine			MC-Granu		
Size of piles (grains)	Time (sec)	Iter	Avg Iter	Time (sec)	Iter	Avg. Iter
25	0.01	56	2.27	0.01	32	14
100	0.10	340	3.41	0.10	221	65
625	0.91	3,191	5.11	6.78	305 213	59 775
2 500	5.03	15834	6.33	-	-	-
5 625	11.46	39206	6.97	-	-	-
8 100	17.61	59308	7.32	-	-	-
10 000	22.29	74352	7.44	-	-	-

4.3 Discussion

The interest of the chosen approach is confirmed by the results obtained and by their validation. As we can see, GranuLab can handle pile-solving problems that couldn't have been considered before.

Moreover, the validation has shown that the model gave a good account of observed macroscopic phenomena that have already been defined in laws (for example, the distribution of forces on the layer in contact with the support). However, we must add that the model considered is too simple to give account of smaller scale phenomena. We are currently working on a more dynamic grain model, which involves disturbing the system at static equilibrium and then studying its dynamic changes. We also still need to better understand the nature of solutions found by the algorithm and their representativeness in the space of all solutions.

If the heuristic “Serpentine” seems to be better adapted to the model considered, the reason for its efficiency is not yet clear. It may be that since the information concerning gravity has to be transmitted downwards before a lower grain in the pile can be solved, this direction is favored. This would explain why, in the model considered of a pile under gravity and not confined, “Serpentine” is the best heuristic. However, the other heuristics could prove useful for other limit conditions as, for example, the case of enclosed piles, under pressure and without gravity (in a horizontal position with a piston on each side).

5 Conclusion

There is often a gap between problem-solving methods used in physics and those developed in Artificial Intelligence. The work presented in this paper shows how a distributed approach can now be implemented and validated to solve complex equation problems that had hitherto remained unsolved. We don't claim that the eco-solving algorithm proposed is the only one that would have solved this problem. However, it is clear that the superiority of this method over the centralized approaches with backtracking comes from the distributed aspects of the problem-solving process and the introduction of heuristics with physical knowledge. In fact, the results show that these heuristics direct search better than a merely stochastic approach would have done. Indeed, in the long term, the objective is to solve moving piles. Through this work, we hope to open up a new path in numerical simulation approaches for granular physics and to contribute to the design and validation of simulations using MAS in physics.

References

1. Collinot, A., Drogoul, A.: La Méthode de Conception Multi-Agent Cassiopée: Application à la Robotique Collective. Rapport LAFORIA n°96/25, 1996, LIP6.
2. Drogoul, A., Dubreuil, C.: A Distributed Approach to N-Puzzle Solving. Distributed Artificial Intelligence Workshop, 1993, USA.
3. Duran, J.: Sables, Poudres et Grains : Introduction à la Physique des Milieux Granulaires. Eyrolles, 1997.

4. Eloy, C., Clément, E.: Stochastic Aspect of the Force Network in a Regular Granular Piling. J. de Phys. I, 7, 1541, 1997, France.
5. Ferber, J., Drogoul, A.: Using Reactive Multi-Agent Systems in Simulation and Problem Solving. Distributed Artificial Intelligence: Theory and Praxis, 1992, p. 53-80, N. M. Avouris and L. Gasser, Bruxelles et Luxembourg, ECSC-EEC-EAEC.
6. Fron, A.: Programmation par Contraintes. Addison-Wesley, 1994, France.
7. Kolb, E., Mazozi, T., Clément, E., Duran, J.: Force Fluctuations in a Vertically Pushed Granular Layer. Eur. Phys. J. B, 8, 483, 1999.
8. Liu, C.-H. et al.: Force Fluctuations in a Bead Pack. Science 269, 1995, p 513.
9. Miller, B., O'Hern, C., Behringer, R.P.: Stress Fluctuations for Continuously Sheared Granular Materials. Phys. Rev. Lett 77, 1996, p. 3110.
10. Ngadi, A., Rajchenbach, J.: Intermitency in the Compression Process of a Granular Medium. Phys. Rev. Lett., 80, 1998, p. 273-276.
11. Ouagenouni, S., Roux, J.-N.: Force Distribution in a Frictionless Granular Pile at the Rigidity Threshold. Europhys. Lett., 39, 1997, p. 117.
12. Radjai, F. et al.: Force Distribution in a Two Dimensional Granular System. Phys. Rev. Lett., 77, 1996, p. 274.
13. Servat, D., Perrier, E., Treuil, J.-P., Drogoul, A.: Towards Virtual Experiments Laboratories : How Multi-Agent Simulations Can Cope with Multiple Scales of Analysis and Viewpoints. Virtual Worlds'98, 1998.
14. Shrager, J., Langley, P.: Computational Approaches to Scientific Discovery. Computational Models of Scientific Discovery and Theory Formation, 1990, Shrager J. & Langley P. (Eds), Morgan Kaufmann.
15. Travers, T., Bideau, D., Gervois, A., Messager, J.C., J. Phys. A, 19, L1033, 1986.
16. Treuil, J.-P., Perrier, E., Cambier, C.: Directions pour une Approche Multi-Agent de la Simulation de Processus Physiques Spatialisés. JFIADSMA'97, 1997, p. 211-228.
17. Valdes-Perez, R.E.: Conjecturing Hidden Entities by Means of Simplicity and Conservation Laws : Machine Discovery in Chemistry. Artificial Intelligence, 65, 1994, p. 247-280.
18. Valdes-Perez, R.E.: Principles of Human Computer Collaboration for Knowledge Discovery in Science. Artificial Intelligence, 107(2), 1999, p. 335-346.
19. Vanel, L., Clément, E.: Pressure Screening and Fluctuations at the Bottom of a Granular Column. Eur. Phys. J. B, 1999, 8, 483.
20. Vanel, L., Howell, D., Clark, J., Behringer, R.P., Clément, E.: Effects of the Construction History on the Stress Distribution under a Sand-Pile. submitted, 1999.

Speeding Up CapBasED-AMS Activities through Multi-Agent Scheduling

Eleanna Kafeza and Kamalakar Karlapalem

Department of Computer Science, Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
{kafeza,kamal}@cs.ust.hk

Abstract. Activity management systems have been employed in large organization to facilitate production time business processes *(or workflows).* An activity consists of interdependent tasks. Exactly one agent (human, hardware/software system) executes a task. CapBasED-AMS is a capability based and event driven activity management system that supports specification and execution of activities using cooperative information systems paradigm. For each business process to be executed an activity instant is created and executed in CapBasED-AMS. Each task of an activity instant is assigned to an agent that executes the task. Therefore, when there are many activity instances executing at the same time many tasks are queued up at agents for execution. In order to speed up the completion of business process, the user can request that certain activity instances be speeded up. In this paper, we design algorithms for speeding up multiple activity instances at the same time by rescheduling activities in multiple agent queues. We evaluate these algorithms for both one-time speed up and multiple times speed up. Our results show that the multiple speed-up algorithm outperform the one-time speed-up algorithm.

1 Introduction

Multiple agent information systems (MAISs) have started to make an impact in supporting next generation on-line Internet driven applications. The main characteristics of these applications are: truly open distributed functionality, flexibility and adaptability, and on-line response to changes in processing environments (automatically or user supported). If a multiple agent system cannot support these characteristics, especially the last one, it will be difficult for it to be a practical and widely adopted solution. Many changes can occur while an application is executed in a MAIS. For example, an agent might be waiting for a response from a failed agent before it can process the next step of solution, thus increasing the response time to the end user. The agents could fail and some of the work done can be lost and must be redone. The goal for solving the problem might change due to external factors (like, change in interest rate). In an Internet driven MAIS, these changes have far reaching consequences and affect the competitiveness of an

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 119-132, 2000.

organization if they are not addressed quickly. Thus, it is pertinent that in a MAIS, there could be requisitions for speeding up some steps of a currently executing application. The objective of this paper is to use workflow systems technology to model this problem and illustrate how MAIS applications modelled as workflows can be speeded up. The applicability of our solution is illustrated within the context of E-Commerce applications. In [11], we address the problem in the context of multiple agents delegating work to other agents.

Agents in WFMS environments are heterogeneous (either humans or software), autonomous in the sense that are responsible for executing the assigned tasks and inter-related, they jointly try for activity achievement and improved performance as a part of a multi-agent distributed system [5]. A lot of work has been done with respect to multi-agents systems (MAS) and their applications [16]. Work on agent co-ordination is presented in [17] but it mainly concentrates on maintaining a diverse set of tasks dependencies over a distributed system. AMS systems in this context have been identified as inflexible due to their lack of adaptation to "on line" changing circumstances [17]. At the same time, the importance of time-related issues in AMS has been identified recently by [6,4] but most research work focus on maintaining task deadlines [10,14,19].

Work has been done on how agents can be used in agent-based marketplaces in order to close a deal [3,7,13]. The authors in [7] present examples of electronic marketplaces where user-agents negotiate the creation of a contract or the purchase of goods. Yet, there is no work describing the next step, that is, after a contract is closed the user might request some additional services on an already agreed upon and executing deal. This is a very common social behavior in traditional markets and it becomes even more important in electronic markets where buyers/sellers deadlines are changing rapidly. In e-commerce environments customer satisfaction has strategic implications therefore there is a need for closely monitoring e-commerce transactions [7]. Hence, besides other functionalities the multi-agent system should be able to respond to events that request the speed-up of already executing instances and apply a speed-up policy to support the request. At the same time as stated in [15] – case study for electronic publishing- most agent-based e-commerce models examine the consumer side. From the e-services provider side, it also beneficial to model and support inter-organization procedures like accessing, order processing, delivery etc in a multi-agent environment. In our approach, we assume that the user agents have already negotiated a contact and they want to alter the agreed completion time. We look into this problem from the service provider point of view and examine how existing executing instances can be speeded-up.

Workflow Management Systems provide the infrastructure to allow business processes to execute in a distributed environment. In the case of B2B, electronic commerce [18] workflow enactment can be supported by a set of collaborating information agents. In an agent-enabled electronic commerce environment agents work cooperatively to solve a diverse set of complex problems, providing the semantics support infrastructure for high-level elements like business processes and workflows. In such environments the authors have identified three different types of agents: *application agents* like goods processing, order handling, *general commerce activity agents*, like negotiation, billing agents and *system-level agents* like distributed workflow managers.

According to the categorization of [18], a collection of programs that provides services to others can be considered as an application agent and the workflow scheduler as a system-level agent. In our context when we refer to agents, we refer to such application agents. Each application agent that is capable of showing speed-up behavior is a *speed-up enabled agent*. Each such agent is autonomous in the sense that it is responsible for deciding whether or not to apply speed-up and select a speed-up policy. The speed-up decision is based on the capabilities of the agent. For example it can decide not to speed-up or it can speed-up only when the number of tasks that it has already accelerated in a given time interval is smaller to a given threshold etc. The workflow scheduler is a system-level agent that propagates information that is crucial for agents' coordination.

There is a lot of work done in agent coordination [17]. In [13] the authors identify the reasons in a multiagent environment that arise the need for coordination as a) to manage dependencies among agent actions b) to adhere and maintain global constraints and c) to facilitate problem solving when one agent cannot solve the problem. In our case, the coordination problem appears when there are several speed-up requests that have to be handled by the agents. In that case, we have a synchronization problem, which is solved through a coordinator agent. In CapBasED-AMS, multi agent workflow environment the workflow scheduler is the coordinator agent. While the objective of each speed-up agent is to execute a set of tasks so that the associated speed-up constraints (if any) are satisfied, the coordinator has an additional global constraint of not slowing-down any activity. Since no additional resources are used for speeding-up an activity, other activities are slowed-down. An unnecessary slow-down occurs when due to poor agents' synchronization, the activity that requested speed-up continues being speeded-up although it has achieved the requested amount of speed-up. In our case, the coordinator agent is responsible for passing the "achieved speed-up" information to the speed-up agents. Speed-up agents independently speed-up the locally residing tasks and they cease to do so when they receive a message that informs them that the activities have achieved their goals. Therefore, we do have some unnecessary slow-down because it is after the current state has completed that speed-up information is collected and propagated to the agents but we have less complexity and more parallelism.

Example: *Suppose that in an e-commerce Internet application a web-based WFMS is used for processing orders. That is, users access the WFMS and specify their requests for buying products. Since at a given moment in time there are several such orders being processed, the WFMS queues tasks for the orders (such as, credit check, approval, personalization, and delivery) are queued up at the agents. Now if a user seeks to speed up his order, then the system should facilitate this speed up. This is a very pertinent problem, because in real life, users do seek speed up of their activities, and the existing systems in order to provide good service, and keep customer loyalty, do speed up the activities.*

In order to deal with such situations, we integrate in CapBasED-AMS [2,9,12] the functionality of speeding up some workflow instances. The core ideas for CapBasED-AMS come from the distributed cooperative problem-solving paradigm, and facilitate specification and execution of activities using a database centric solution [2]. The idea behind speed-up is to achieve a faster execution of some instances *without acquiring more resources*, therefore without increasing the cost of the execution. There is always an overhead to be paid; some other activity instances are going to be slowed down. Yet, we increase the possibility that urgent tasks, whose delay might have severe consequences, will finish earlier.

The contribution of the paper is the development of a scheduling framework for speeding up activities which incorporates:

- A model of a multi-agent speed-up environment (location table model) and of the speed-up problem,
- Speed-up scheduling policies (one-shot and multiple solution),
- Evaluation of the utility of the algorithms using simulation

The paper is organized as follows: Section 2 is the CapBaED-AMS architecture for speeding up activities' execution. In section 3, we formulate the speed-up problem for activity instant executions. Scheduling is presented in section 4. The simulation model and the experimental results are presented in section 5.

2 ↑↓CapBasED-AMS

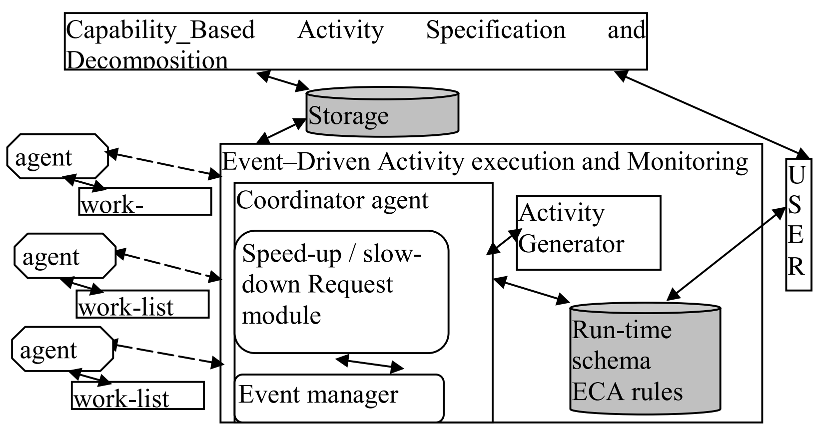


Fig. 1. The ↑↓CapBasED-AMS Architecture

As presented in [2,9,12] CapBasED architecture consists of two parts 1) the capability based activity specification and decomposition, and 2) the Event-driven activity execution and monitoring. In this paper, we address only issues regarding the integration of the speed-up/slow-down capability in the system. In the top part (Figure 1) of the architecture an activity is specified and decomposed into its constitute tasks. A special module the Match Maker is responsible for identifying for every task a corresponding capable agent. This matching process is based on tokens: tasks that need the speed-up capability and agents that possess the capability of speeding up the task execution have an attached token: (*token_id*, *description*)=(SU, "able to speed-up"). Further details about the CapBasED-AMS approach to Match Making can be found in [9].

In the bottom part of the architecture, the Activity Generator Module creates the activity graph, ECA rules, the runtime database schema, or the nested activity and its corresponding metadata. The Coordinator agent is the execution manager responsible for scheduling and synchronizing the execution of tasks of activity instances. It interprets the activity graph, determines tasks that are ready for execution or submits the sub-activities and speed-up information to the application agents. The Event manager is responsible for handling the events.

Each agent is an *application agent* that is an agent responsible for executing a task like credit test, billing services, product delivery etc. Each application agent is *autonomous* i.e. capable of executing without human intervention when it is a system agent, it has a *social ability* interacts constructively with other agents, it can *respond* in a timely fashion to events that are submitted by the event manager and can initiate decision when the situations demand it (*pro-activity*). Hence, agents in ↑↓CapBasED-AMS have the agent-required attributes as described in [8], and can execute and deliver the necessary e-commerce services. In addition, they have the speed-up capability. That is, in a system failure or upon an explicit user request they initiate speed-up procedures in order to handle the situation.

3 The Location Table Model

3.1 The Location Table

Work-lists indicate at a given point in time the tasks that are ready to execute by the agent. Each agent takes the first task from its corresponding work-list and executes it. Each work-list is represented as a queue. The following matrix describes the state of the work-lists at any given point in time:

Definition: The *Location Table* (LT) is a $p \times n$ matrix where n is number of activity instances that have submitted a task to the agents and p is the number of agents. Each row of the location table represents an agent queue. The elements of the table (x_{ij}) are tuples $<a_k,t_l>$ representing that activity a_k has submitted task t_l for execution and that task t_l occupies j^{th} queue position in agent i.

We can view the location table as a snapshot of the agent's environment. The location table is a *dynamic* data structure routinely updated upon completion of a task by an agent according to FCFS policy and when a speed up request is processed.

Example 3.1: *Suppose the following process specification:* A=AND(SEQ(t_1,t_2,t_3), t_4,SEQ(t_5,t_6), t_7). *A pictorial representation of activity process A is given in the Figure 2, where ANDs denotes the split of the AND construct and ANDj the join of its branches.*

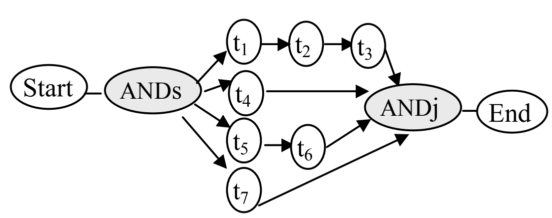


Fig. 2. Activity process specification

Let A_1, A_2, A_3, A_4, A_5 and A_6 be six active instances of an activity process A. Upon start of execution there are 24 tasks enabled. This is the set of tasks that fulfill the precedence constraints -four tasks from each instance (A_i,{t_1,t_4,t_5,t_7}).Figure 3 shows a possible location table. The first cell $<A_1,t_1>$ indicates that activity process instance A_1 has submitted task t_1 for execution to agent$_1$ and that t_1 occupies position 0 of the agent's queue (the first queue position).

pos	*0*	*1*	*2*	*3*	*4*	*5*
ag_1	A_1,t_1	A_2,t_1	A_3,t_1	A_6,t_1	A_5,t_1	A_4,t_1
ag_2	A_4,t_4	A_1,t_5	A_2,t_4	A_2,t_5	A_3,t_4	A_1,t_4
ag_3	A_5,t_4	A_6,t_4	A_3,t_5	A_4,t_5	A_1,t_7	A_2,t_7
ag_4	A_5,t_5	A_6,t_6	A_3,t_7	A_5,t_7	A_4,t_7	A_6,t_7

Fig. 3. Location table

3.2 Speed-Up by Positional Values

We introduce a measure on the expected execution time for a given snapshot by associating with every activity instance task a *positional value*. The positional value is a relative measure; it does not indicate actual time, it gives a measure on the queue positions that the activity's tasks have to go through, before acquiring the agent.

Definition: Given an activity instance A_k and a location table LT, we define *the positional value of element* $<A_k,t_l>$ *denoted as* $pv_{LT}(<A_k,t_l>)$ as the number of the tasks that precede it in the Location table row that represents the agent queue the element resides.

The elements' positional value is a number that denotes its queue position under the condition that the first element in the queue is assigned position 0. Semantically the positional value denotes how many tasks the element $<A_k,t_l>$ has to wait before it acquires the first queue position thus before being the next task to be executed. We use the notation $t_i < t_j$ to denote that task t_i and t_j reside on the same *agent and* t_i *proceeds in agents' queue* t_j.

Definition: Given an activity instance A_k and a location table state LT, the *positional value of instance* A_k is the following sum:

$$pv_{LT}(A_k) = \sum_{\forall t_j, t_n \in A_k} \left(pv_{LT}(\langle A_k, t_j \rangle) - \sum_{\forall t_n:\ t_n < t_j} pv_{LT}(\langle A_k, t_n \rangle) \right)$$

The positional value of an activity instance is the summation, for all agents, of all the tasks' positions (i.e. tasks' positional values) of the given activity instance. The summation represents in total how many positions A_k has to wait before all its tasks currently in the queues have started or completed execution. If an activity has submitted more than one task to a given agent then we deduct the positions added by the previous tasks because it duplicates the number of tasks (positions) the activity has to wait before the last task in the current agent starts executing.

Given an LT the set of all its positional values form its *positional vector* PV. The positional value helps the user to find out the relative order in which activity instances complete their executions, and seek speed up of some of the activity instances.

Example 3.2: *Consider the location table of the previous example (figure 3). The queue positions are depicted in bold font. The positional value of every element in each column is the POS value of the column. For example tuples* $<A_1,t_1>$, $<A_4,t_4>$, $<A_5,t_4>$ *and* $<A_5,t_5>$ *have positional value 0 because they reside in column 0 of the location table. The positional value of instance* A_1 $pv(A_1)=pv(<A_1,t_1>)+pv(<A_1,t_5>)+pv(<A_1,t_4>)+pv(<A_1,t)-pv(<A_1,t_5>)=0+1+5+4-1=9$ *(the cells are depicted in gray colour in LT). Note that in agent*$_2$ *instance* A_1 *has submitted two tasks* t_5 *and* t_4 *but the instance totally has to wait for* $pv(<A_1,t_4>)=4$ *tasks in agent*$_2$*. Therefore, we have to* *deduce the positional value of* t_5 (*1*) *from the summation of the positional values. The Positional Vector of LT is PV= (0+5+4, 1+3+5, 2+4+2+2, 5+3+4, 4+3, 3+1+5)= (9,9,10,12,7,9) where each element corresponds to instance* A_1 *to* A_6*, respectively.*

Definition: A location table LT′ is an *alias* for location table LT iff LT′ results from LT by permuting the tuples of one or more rows.

Each such permutation is actually a re-arrangement of the tasks for a specific agent queue. We next define *snapshot speed-up*. Similarly, we can define *snapshot slow down* as well.

Definition: Let LT be the $n \times p$ location table for a given snapshot. We say that location table LT′ achieves a *snapshot speed-up* for activity instance A_i with respect to LT iff LT′ is an *alias* for LT, and $pv'(A_i) < pv(A_i)$

Example 3.3: *Let LT′ be an alias of LT presented in example3.2 as follows:*

pos	*0*	*1*	*2*	*3*	*4*	*5*
ag_1	A_1,t_1	A_2,t_1	A_3,t_1	A_4,t_1	A_5,t_1	A_6,t_1
ag_2	A_4,t_4	A_1,t_5	A_2,t_4	A_2,t_5	A_1,t_4	A_3,t_4
ag_3	A_2,t_7	A_1,t_7	A_4,t_5	A_3,t_5	A_6,t_4	A_5,t_4
ag_4	A_4,t_7	A_6,t_6	A_3,t_7	A_5,t_7	A_5,t_5	A_6,t_7

The positional vector of LT′ is PV′=(5,4,12,5,13,14). Activities A_1, A_2 *and* A_4 *are speeded up, as 5<9, 4<9 and 5<12 for* A_1, A_2 *and* A_4*, respectively.*

In each location table for every activity there is a bound on the amount of speed-up it can achieve. In the best case, all activity's tasks will occupy the first queue positions. Therefore, the *max-speed-up value* an activity can achieve equals the summation of the positions of the tasks that are ahead of it in the queue (which is the speed-up gained if it is moved to the head of the queue). For example, in LT the *max-speed-up value* of A_2 is 9. We see that an activity's max-speed-up is its positional value.

Definition: Given a location table LT and an alias LT′ we say that activity instance A_i has achieved its *max-speed-up* iff $pd_i = pv_i(A_i)$, where $pv_i(A_i)$ is A_i*'s positional value* and pd_i is A_i*'s positional- difference value*.

We use positional vectors for identifying and measuring speed-up. By comparing the PVs of two tables where one is the alias of another, we can determine which instances were speeded-up, slowed-down or remained unchanged and the amount of speed-up/slowed-down achieved.

Definition: Given two location tables LT and LT′ and their corresponding positional vectors PV and PV′, we define as their *positional-difference vector* $PD_{PV,PV'}$, the vector $PD_{PV,PV'}$=PV-PV′.

A positive positional-difference value, say k, denotes that the instance has been accelerated k positions while a negative value (−k) denotes that the instance has been slowed-down k positions.

Example 3.4: *In example 3.3 the positional–difference vector is* $PD_{PV,PV'}$*= (9,9,10,12,7,9) −(5,4,12,5,13,14)= (4,5,−2,7,−6,−5). Activities* A_1,A_2,A_4 *were speeded-up because the positional-difference values are positive, by 4,5 and 7 positions respectively. Activities* A_3, A_5 *and* A_6 *were slowed-down because the positional-difference values are negative by 2,6 and 5 positions respectively.*

4 Scheduling

The activities are speeded up by using agent queues. The speed-up/slow-down problem in our framework is translated as follows:

Let LT be a location table and let RV be a *request vector* such that $\forall rv_i \in \mathrm{RV}$ A_i requested a speed-up of rv_i time units (if rv_i is 0 A_i requested no speed-up) such that $\forall rv_i \in \mathrm{RV}$, $rv_i < pv(A_i)$. Find an alias LT′ such that $\forall rv_i \in \mathrm{RV}$ and $rv_i > 0$, $pd_i \geq rv_i$.

This is a search problem where in the exhaustive approach we have to start producing all possible aliases of each agents' queue (each row of the location table) and stop when the RV is achieved. In the worst case, we have to produce all possible location tables. That is, $(n!)^p$ location tables for a $n \times p$ location table. Clearly, this is an infeasible solution for on line scheduling. In our work, we propose a polynomial time heuristic tailored to the needs of the multi-agent speed-up problem.

We propose two different modes of execution: *one-shot* and *multiple speed-up mode*. The *one-shot* policy tries to achieve the requested speed-up in the current snapshot while the *multiple speed-up* policy tries to achieve a portion of the requested speed-up over multiple location tables. The one-shot policy tries to achieve as much speed up possible as soon as possible. Whereas multiple speed-up policy tries to achieve the required speedup as much as possible over time until the activity completes its execution. Thus, they cater to different types of speed-up requests. The multiple speed-up policy always satisfies better the user request but it needs a longer time period for satisfying the user request. One-shot policy is useful in the extreme scenarios wherein the multiple step policy cannot be applied.

We present one algorithm for each policy: the *One-shot Maximum Position First* (OMPF) and the *Multiple Maximum Position First* (MMPF), respectively. Both speed-up-scheduling algorithms execute in a *multi-agent environment* and are *agent independent*. Each agent speed-ups the tasks residing at his site *independently* of other agents' speed-up based solely on the information attached to the submitted tasks. We present the structure and the components of the multi-agent environment in section 6.

4.1 The One-Shot Maximum Position First

The *One shot Maximum-Position-First algorithm* (OMPF) heuristic tries to push forward on the queue all the tasks that have requested speed-up. It does so by putting in the first queue positions those tasks that are further behind in the queue (figure 4).

In the One-shot Maximum Position First (OMPF) algorithm the swapping decision is based on the elements positions in the queue. For every agent queue, it accelerates activity instances that have the greatest position. From those activities that requested a speed-up, the ones that are in the back of the queue are served first, and they are pushed as much forward as possible. This policy tries to swap "small" positions with "large" ones. If the swap is more than the requested speed up it still swaps and achieves more than speed-up than requested.

Input: LT[i], Request Vector
Output: Positional Difference vector
Maximum-Position-First algorithm (OMPF) {
  For the agents' queue:
    Lock all instances that have requested speed-up and should not be slowed down
    Sort in descending order of position the instances requested speed-up and they have not still achieved it and are not locked //SU vector
    Sort in ascending order of position the instances that have not requested a speed-up //SD vector
    Swap every speed-up task (SU) with the first slow-down task (SD) that has smaller position and is not locked.
    Calculate the remaining speed-up;
  return PD;
}

Fig. 4. The OMPF algorithm

The rational behind the OMPF algorithm is to achieve as much speed-up as possible for all activities that requested speed-up. By swapping each time the elements with the maximum distance in the queue, it achieves totally for the set of all activities that requested speed-up the maximum possible speed-up, in every agent. The drawback is that in doing so in most cases activities are speeded-up more than requested forcing the rest ones to slow down unnecessarily.

Example 4.1: *We assume the location table of example 3.1 and let the request vector be RV=(7,5,0,7,0,0), that is activity instances* A_1,A_2 *and* A_4 *requested speed-up.*

The SU and SD vectors for the $agent_1$ *based on LT are (5,1,0) and (2,3,4) respectively. The swapped pair is (5,2) where task* $<A_4,t_1>$ *in position 5 of LT is swapped with task* $<A_3,t_1>$ *in position 2. For* $agent_2$ *SU=(5,3,2,1,0) and SD=(4). The swapped pair is (5,4) with corresponding tasks* $<A_1,t_4>$ *and* $<A_3,t_4>$*. Similarly for the rest of the agents. The LT alias* LT^{OMPF} *of LT generated by the OMPF algorithm is shown below.*

POS:	**0**	**1**	**2**	**3**	**4**	**5**
ag_1	A_1,t_1	A_2,t_1	A_4,t_1	A_6,t_1	A_5,t_1	A_3,t_1
ag_2	A_4,t_4	A_1,t_5	A_2,t_4	A_2,t_5	A_1,t_4	A_3,t_4
ag_3	A_2,t_7	A_1,t_7	A_4,t_5	A_3,t_5	A_6,t_4	A_5,t_4
ag_4	A_4,t_7	A_6,t_6	A_3,t_7	A_5,t_7	A_5,t_5	A_6,t_7

The new alias positional-vector is PV^{OMPF}*=(5,4,15,4,13,12). The PD= (4,5,–5,8, –6,–3). We see that* A_1 *did not achieve its requested speed-up (it requested 7 and achieved 4),* A_2 *achieved the exact requested value while* A_4 *achieved more than requested.*

The complexity of the algorithm is O(*n*∗*logn*) for every agent , where n is the number of elements in the queue. For every iteration, we need O(*n*) accesses for creating the speed-up and slow-down vectors, where *n* is the number of instances. We need *O(n logn)* time to sort the two vectors. In the swapping phase in the worst case each element of both vectors is accessed only once, therefore we need to add O(*n*) time for the swapping phase.

4.2 Multiple Speed-Up Solution

While the OMPF algorithm tries to achieve the speed-up request vector by taking a snapshot of the system, the current location table, the Multiple Maximum Position First (MMPF) algorithm tries to achieve the speed-up goal over several snapshots.

At a given point in time, when the speed-up request is posed there are several active instances executing. The MMPF algorithm first applies the OMPF on the first location table. That is every agent independently speed-up its queue and return the achieved speed-up to the AMS. After all tasks of the current location table have completed execution, a new location table is created. It contains the new set of enabled tasks i.e. tasks submitted for execution by the precedence constraints scheduler of the AMS so that all the precedence constraints are satisfied. The MMPF scheduler in each agent for every task checks whether the requested speed-up has been achieved. If yes, it locks the tasks so that the activity will not be slow-down, if not it reapplies the OMPF on the new location table. The algorithm iterates until either all activities achieve their requested speed-up or there are no more tasks in the activities to speed-up.

The one-shot algorithm can be viewed as a special case of the multiple-times one when the number of location tables is set to one.

5 Experimental Evaluation

5.1 Activity Process Multi-agent Simulator

In our experimental evaluation we simulated a multi-agent environment were several inter-dependent activity instances are executing at the same time. The simulator is composed by: the *activity coordination* module and the *Agents*. The *activity coordinator module* is responsible for submitting tasks according to their temporal precedence constraints, attaching to tasks the necessary speed-up information and it evaluates the results returned by the agents. *Agents* re-schedule tasks according to the algorithm and report the achieved results (figure 5).

The *Activity process generator* generates randomly activity processes. The *Activity instance generator* monitors the execution of each single instance. The user initially specifies the number of active instances. The *Speed-up request generator* upon user request generates a set of speed-up activity requests, also keeps track of the corresponding amount of speed-up requested by each instance.

The *Precedence-Constraint scheduler* is responsible for searching among all active instances for those tasks whose precedence constraints are satisfied. It finds the tasks that all its predecessors have been already scheduled and completed. It sends those tasks to the Enabled Tasks scheduler with attached information regarding the identification of the activity process and the task identification. The *Enabled Task scheduler* attaches to every enabled task already sent by the Precedence-Constraint scheduler, speed-up information and submits the task for execution to the corresponding agent. The *Results evaluation module* accepts agent's responses and keeps track of the activities' speed-up achieved values. By the end of the execution, it evaluates the results and sends statistical information to the user. *Agents* are queues that hold tasks and a scheduler that arranges the order of task execution. The speed-up algorithm resides at the agent site and executes independently from the centralized system based on the information attached to every submitted task.

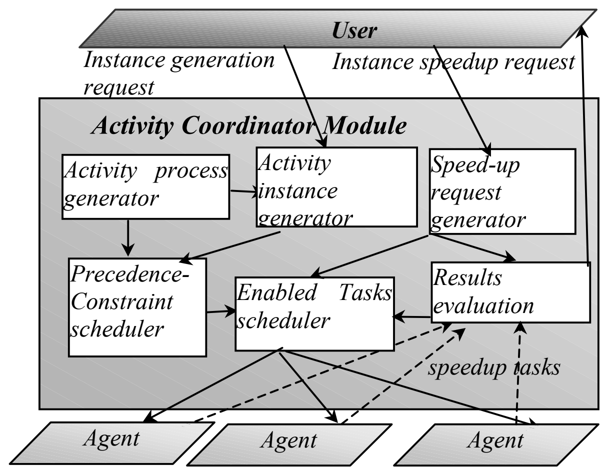


Fig. 5. The Multi-agent simulator

We assume that the agents use FCFS policy and that the user imposes the speed-up requests at the moment of the first location table creation and until all currently executing instances finish, no new speed-up request is posed. These assumptions simplify the setting of the parameters and facilitate the monitoring of the execution focusing on speed-up.

5.2 Parameters Set-Up

Several parameters have to be set in the environment. The following table shows a brief description of each parameter and the value or the range of vales that we used in the simulation.

Parameters	Default Values
Number of agents	10
Number of processes	1-4
Number of instances/ process	500-2000
Total number of intances	2000
Max Number of tasks/ instance	50
Number of activites in LT requested speedup	{0.1,0.3,0.5,0.7}*number of activities in LT
Amount of speed-up requested	{0.1– 1}*max-speed-up

We calculate the amount of requested speed-up as a portion of the *max-speed-up* of the activity instance of the snapshot LT. For every activity instance A_i, the rv_i varies from 0.1**max-speed-up to max-speed-up* of A_i. As performance measure, we use *achieved-ratio* (AR): the number of instances that achieved the requested speed-up versus the total number of activities that requested speed-up. An AR of 1.0 (100%) implies that the algorithm has satisfied all speed-up requests. For our simulation study, we used C++ on a 400 MHz Pentium II Windows NT workstation.

5.3 Results

We conducted the experiments for 2.000 activity instances. The x-axis determines the requested speed-up and y-axis represents the AR. In each figure, there are five plots, each one for a different number (1,3,5,all) of location tables. For Figures 7 and 8, we see that as long as the amount of requested speed-up is less than 0.3**max-speed-up,* for all algorithms, most activities achieve their requested speed-up. As the amount of requested speed-up is in excess of 70%, more location tables are needed for MMPF in order to achieve the requested speed-up. For OMPF the success radically declines reaching zero for *max-speed-up.* In the best case, only one instance that occupies all 0-queue positions achieves *max-speed-up*. The only feasible value for the *achieve-ratio* when *max-speed-up* is requested is 1/*number of instances that requested speed-up*. Therefore, for one location table, which is the case of the OMPF, we expect all algorithms to end-up to 0 as the requested speed-up reaches *max-speed-up*.

As the number of instances to speed-up increases (Figures 9 and 10) for OMPF, less activities manage to achieve smaller percentages. This is because the number of activities that can be slowed-down decreases as well therefore the number of available queue positions for swapping in the LT decrease as well. We observe that the multiple-shot algorithm MMPF always achieve the requested speed-up. This is because there are several chances in subsequent location tables for tasks to be speeded-up. We can conclude from the dashed-line plots that when the number of activities that requested speed-up increase the number of location tables needed to achieve the request, increases as well.

When activities are executed in multi-agent environment, it is very much feasible to speed up some of the activities based on user requirements. Such facilities in an multi-agent problem solving environments is very much needed, as one is not sure when certain activities need to be speeded up.

6 Conclusions and Further Work

In this paper, we study the problem of speeding up activities in a multi-agent environment. We addressed the problem of speeding-up activity instances executions in a multi-agent environment without acquiring more resources, while agents serve the speed-up requests independently. We have defined a positional metric for counting speed-up/slow-down and proposed a heuristic algorithm for achieving the requested speed-up. Speed-up is achieved by re-scheduling tasks positions in agent's queues. We introduced a measurement of speed-up, which takes into account tasks' positions in the agent queues. Our experiments suggest that multiple step solution behave better than one-time speed-up efforts.

Issues such as, development of remedies for those instances that were slowed-down, and the development of a decision-making policy for choosing whether slow-down is more profitable than speed-up need to be addressed.

References

1. Alonso G., Hagen C., and Lazcano A.: Processes in Electronic Commerce. ICDCS Workshop on E-Commerce and Web-Based Applications, 1999.
2. Chakravarthy S., Karlapalem K., Navathe S. B. and Tanaka A., Database Supported Cooperative Problem Solving, *Int. Jour. of Intelligent and CIS,* 2,3 248-287, 1993.
3. Chavez A., Maes P., Kasbah: An Agent Marketplace for Buying and selling Goods, *Proceeding of PAAM 1996*.
4. Dadam P., Reichert M., The ADEPT WfMS project at the university of Ulm, *WPM'98*.
5. Dogac A., Kalinichenko L., Ozsu M.T., Sheth A.. Advances in WFMS and Interoperability. *NATO ASI* 1997.
6. Eder J., Panagos E., Rabinovich M., Time Constraints in Workflow Systems, *CAiSE 99*.
7. Eriksonn J., Finee N.,.Janson S, SICS MarketSpace-An Agent-Based Market Infrastructure, Agent Mediated Electronic Commerce (*AMET 9*8) Lecture Notes in Artificial Intelligence, vol 1571, 1998.
8. Guttman R., Maes A. P., Agent-mediated Interactive Negotiation for Retail Electronic Commerce, Workshop on Agent Mediated Electronic Trading (*AMET 1998*) 1998.
9. Hung P. C. K., Yeung H. P. and Karlapalem K., *CapBasED-AMS*: A Capability-based and Event-driven Activity Management System, in *ACM SIGMOD*, 1996.
10. InConcert. Technical product overview. Xsoft. a division of Xerox. 3400 Hillview Avenue, Palo Alto.
11. Kafeza E. and Karlapalem K., Speeding up CapBasED Activities in Delegation Driven Multi-Agent Information Systems, to appear in AOIS-2000.
12. Karlapalem K., Yeung H. P., and Hung C. K., *CapBasED-AMS:* A Framework for Capability Based and Event Driven Activity Management System, in *CoopIS'95*, 1995.
13. Kerschberg L., S. Banerjee S., An Agency-based Framework for Electronic Business, (*CIA'99*) Cooperative Information Agents III, Lecture Notes in Artificial Intelligence, vol 1652,1999.
14. Leymann F.and Roller D.. Business process management with flowmark. *Proceedings of the 39th IEEE Computer Society International Conference, p:231-233,* California, February 1994.
15. Muller J.P., Pischel M., Doing Business in the Information Marketplace-A case study, *Proceeding of Autonomous Agents 1999*, Seattle 1999
16. *Multi-Agent Systems and Agent-Based Simulation*, MABS '98 Paris, France, 1998.
17. Ossowski S., *Co-ordination in Artificial Agent Societies*, Social Structures and Its Implications for Autonomous Problem Solving Agents, Springer-Verlag, 1999.

18. Papazoglou M.P., The Role of Agent Technology in Business to Business Electronic Commerce, (*CIA '99*) Cooperative Information Agents III, Lecture Notes in Artificial Intelligence, vol 1652,1999.
19. TeamWare Flow. Collaborate workflow system for the way people work. P.O.Box 780,Helsinki, Finland.

Appendix

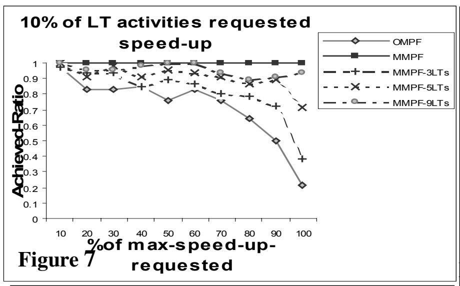


Figure 7

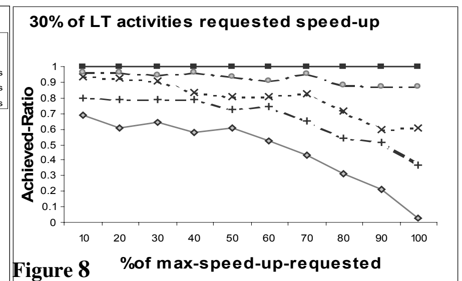


Figure 8

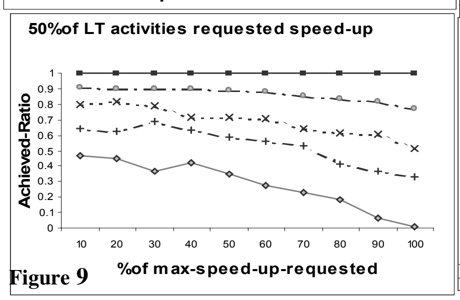


Figure 9

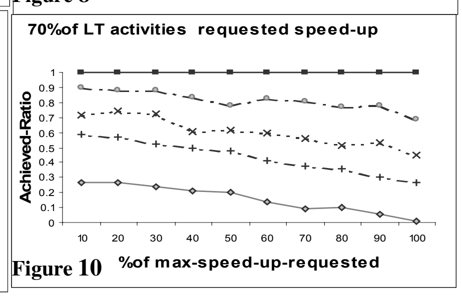


Figure 10

The Micro-Macro Link in DAI and Sociology

Michael Schillo[1], Klaus Fischer[2], and Christof T. Klein[3]

[1]Multi-Agent Systems Group, Saarland University,
Im Stadtwald, 66123 Saarbrücken, Germany
schillo@virtosphere.de
[2]DFKI GmbH,
Stuhlsatzenweg 3, 66123 Saarbrücken,
kuf@dfki.de
[3]Department of Sociology, Saarland University,
Im Stadtwald, 66123 Saarbrücken, Germany
ctk@ags.uni-sb.de

Abstract. No matter if a population is human or artificial, we can surely identify phenomena that can be described as micro or macro phenomena. In this paper, we discuss micro and macro aspects of a population from a DAI and a sociological point of view. We analyse similarities and differences in these viewpoints, and identify misperceptions in the DAI community about the micro-macro terminology. We explain these misperceptions and argue for the transfer of sociologically founded concepts to agent-based social simulation. Our research is done in the DFG focus programme socionics. We cooperate with sociologists from University Hamburg-Harburg with the intention to transfer knowledge from sociology to DAI as well as from DAI to sociology. In cooperation with DFKI Saarbrücken we work on improving agent theories to be applied in large sized multi-agent systems in the freight logistics domain.*

1 Introduction

The problem of how individual action and structural rules in a set of agents interact is a foundational issue for both DAI and sociology, also known as the micro-macro problem. The understanding of the link between micro and macro would mean a substantial advance in designing agents for dynamic and large-scale agent-based social simulation, as well as a deeper understanding of human societies. Furthermore, modelling the macro aspect in agent theories is considered to be essential for DAI research, as this concept substantially contributes to the distinction between artificial intelligence and distributed artificial intelligence (DAI). For this enterprise, a scientific cooperation with sociology can be of great benefit to DAI. However, we found that a mutually agreed terminology cannot be assumed.

The micro-macro problem is perceived in distributed artificial intelligence (DAI) research as a central issue because it directly refers to such problems as coordination and scalability. And indeed, the definition of distributed AI as opposed to the parent

* This work is supported by Deutsche Forschungsgemeinschaft under contract Fi 420/1-1.

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 133-148, 2000.

discipline of artificial intelligence heavily depends on aspects that are only introduced by the problems that occur when multiple actors face the results of each other's actions [47]. Not surprisingly, there are differences of definitions of the micro-macro problem as researchers perceive it in the DAI community and the perspective taken in mainstream sociology. DAI definitions of the macro level either intend to abstract from the individual and to summarise certain features of a group of agents (performance, communication overhead etc.) or aim at mechanism and organisational design. While the former is a descriptive approach, the later is normative. Sociology does study the same level of abstraction from the individual, but takes a different (and usually only descriptive) perspective. In sociology the macro level of a society is itself a structure, which possesses to a certain degree it's own autonomy: it survives the individual and is (primarily) independent from the influence of any single individual. A further important feature of the macro level is that it reproduces itself over and over again by channelling the interests of the individuals.

While organisational theory by definition does not make any claims about how a society (including a number of organisations) is composed, reducing the complexity of a society to a multi-dimensional performance vector does not pay tribute to the complex dynamics that can be observed at the macro level of human societies. This does not render the cooperation of the two fields obsolete. On the contrary, looking further at sociological theory is most fruitful to DAI research. Apart from the solely action-oriented or structure-oriented theories, there is a selection of hybrid theories that try to explain the connection between individual action and social structure (Giddens, Bourdieu etc.).

In the discussion section we propose and start to analyse the habitus-field theory of Pierre Bourdieu, which tries to explain the effect of individual behaviour on societal structures *and* vice versa. This is where the great strength of the theory lies and where we expect that DAI will find a lot of concepts for overcoming the micro/macro gap. For example we state that the theory on this reciprocal relationship is the medium that answers Castelfranchi and Conte's [7] question of how cognition can be structured by society and what is essential for the emergence of structure from micro-interactions.

Our research is done in the context of the field of *socionics* [34]. In this area sociologists and computer scientists try to transfer methods and theories from one discipline to the other. Our main concern is the modelling of interactions in the domain of shipping companies. This scenario is defined by its openness and complexity as we encounter a great diversity of agents as well as tasks and time restrictions. Typically is also the large scale of such MAS in the magnitude of thousand agents that requires not only interaction on a micro level but also macro structures to function efficiently, and coherently. Our work leads us to the conclusion that building social agent architectures that can deal with both, micro and macro phenomena is not solely for the purpose of human adequacy but has also strict engineering reasons. This emphasises the importance of sociologically founded theories applied in DAI research.

2 The Micro Level in DAI and Sociology

The micro level is the area where we can expect to find mostly agreement between the two disciplines. The micro level is composed of individual actors (humans and agents, respectively) that interact. However, both disciplines emphasise different aspects. DAI focuses on the cognitive architecture and the theory of how to model knowledge acquisition and memory, perception and problem solving. This results in a focus on designing algorithms that produce for a given input an appropriate (rational?) output, as expressed by the widespread acceptance of decision and game theory. Sociology on the micro level however, focuses on *inter*action and relationships between actions and actors. Also, sociologists consider social actions, i.e. actions that are aimed at changing the actions, effects of actions or beliefs of another individual. It is important to note that this excludes actions like unwillingly causing an effect on another person and actions aimed at objects, but includes actions like threatening another person (social in the sense of related to other individuals and not in the sense of caring). These differences may seem subtle at first. As we go on to take a look at the macro level, this differences become more important, as the perception of *what* is interacting on the macro level diverge significantly.

3 Overview on Perspectives on the Micro-Macro Link in Sociology

An exhaustive discussion of the definition of the macro concept in sociology definitely exceeds the space provided here (and our competence). In fact, this discussion fills volumes and some will even argue that this discussion is equivalent to doing sociological research. We can note that many definitions exist, all tailored to a specific theory, with no apparent success in the discipline to generalise from specific theories. A second problem with presenting a clear-cut definition of the sociological notion of the macro level may be that there is no corresponding physical fact in reality. Even the phenomena usually connected with certain levels (e.g. interaction for the micro level) are hard to pin down as they sometimes are used with slightly differing connotations (e.g. when talking about the interaction of religion and politics as their bi-directional influences, which are phenomena of the macro level).

Depending upon perspective of observation, the subject of social sciences can be examined thereafter similarly from micro, meso, to macro or metasociological perspective. The missing of a generally accepted theory of the social leads to distortions and formation of different schools with according to differentiated research programs. Thus different paradigms co-exist for the study of the emergence of social structure in contrast to Kuhn's thesis on „changes of paradigms" [26]. A brief description of the four perspectives follows:

Micro-level: Sociology as science of social concern and interhuman behaviour. Investigation of the influences of small groups on the non-standard behaviour (concern, perception and thinking), e.g. groups and exchange theories.

Meso-level: Sociology as science of the social institutions and organisations. Investigation of the influences from social organisations, e.g.: organisation sociology, work sociology, technique sociology, sociology of education.

Macro-level: Sociology as science of the whole society, its stability (static aspects) and change (dynamical aspects). It analyses which forces are responsible for stability and change: religion, economics, culture, institutions etc. Investigates the influences of the 'society' and culture, e.g. general system theory, sociology of culture.

Meta-level: Sociology as science of the ideas about society and as criticism of ideology. Investigates society and culture constructing ideas, objects and values, e.g. knowledge sociology, social philosophy, critical society theory (Frankfurter Schule).

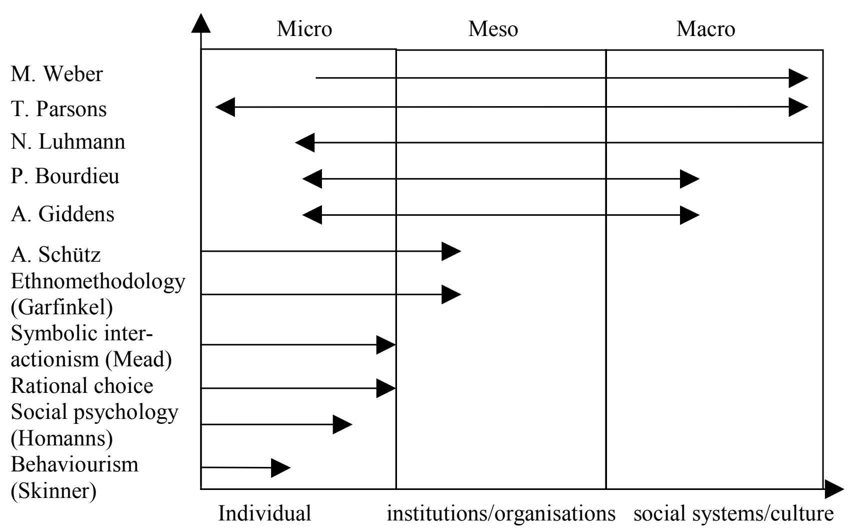


Fig.1 Overview on foundational strands in sociological theories

For example the introduction of a bank holiday will surely provoke a wide discussion in a society. Politicians, trade unions, employer associations, even the churches will engage in a debate on the advantages and disadvantages, all parties with their respective motives. Observing the influences of the different fields that interact here (economics, politics, religion) is observation of the society on the *macro* level. A *meso* level view would be e.g. the investigation of the different groups involved. Maybe a new movement will form that aims to prevent this bank holiday. An investigation of this movement would be a meso study. A *micro* level observation would be, if we looked at individuals in a group confronted with this topic and how they interact, which group processes exist that shape the interaction etc. A *meta* observation could be how eastern societies and western societies differ in decision making on topics that involve economics, religion and politics.

When evaluating current sociological theory, we need to take into account the classical theoretical works in this discipline. The literature on social theory presents itself as complex and multi-layered. The social life as the shared object of investigation was re-built as a complex variation of phenomena, depending on observation levels by the examining scientists and their specific ways of examination. In order to classify the parts of the social universe, it was broken up into four levels as described above.

In a short overview we will present a collection of social theories and briefly discuss their ranges and main features from the perspective of DAI. We divide the broad range of theories into theories focusing on the micro level, focusing on the macro level and theories that try to translate from the micro to the macro level and back (see Figure 1). As we focus on the micro-macro link, we leave the more abstract meta level out of scope of this discussion. Of course, this overview is reducing the theories to an absolute minimum and will by many (sociologists) be viewed as lacking respect for the complexity of the theories. However, this overview is not intended to cover the theories in their details, this would be impossible in the space given to a paper and a complete meta-analysis is left to scientists with more competence.

3.1 The Macro-Approaches

In the centre of these approaches lie large social formations or collective processes (the objective structure). Their objects are for example the structure and the change of governmental organisations and institutions (e.g. capitalist society formations as strata, classes, parties). The main interest is to attempt the analysis of the whole society by its objectified social structures. The aim of this macro-orientation on social life phenomena is to describe and explain processes of reproduction (static aspect) and social change (dynamic aspects) of societies under economical, social and cultural points of view. The society is to be considered as an reality of its own, which can not be deduced from individual contexts (i.e.. from acting and behaviour). In this view the society does not comply with the sum of its parts.

The individual subjects play a minor part for the constitution of social life and its actual conditions. In fact, by reconstructing the social in macro-models their influence on social structure merely occurs as exchangeable data (contingent functions of individuals). See for example the so called *normative paradigm* of Parsons' action-theory [36]: Confronted with social expectations (may-, shall-, must-expectations), the owner of a social position (objective social structure) will, in spite of Parsons voluntarist assumptions, be forced under societal conditions with different degrees of sanctions to adapt to the objective structure. Thus, the *homo sociologicus* is viewed as fulfilling obediently the integrative and forced upon function of the more abstract layer in the social system. To give an example: In Luhmann's conception of social systems [28][29] the actors were completely excluded with the definition of communication as the basic element of modern societies and the selective process of *information, mediation* and *understanding*, his theory of social systems defines the individual (*psychical systems*) as *environment to the system*, which can only participate to the social by communication. (cf. [36, 38, 1, 2, 46]).

3.2 The Micro Approaches

The micro-sociological approaches study the social by observing the individuals and their interaction behaviour (e.g. [17]). The issue most important in this research area is: How can individual behaviour (action, mind, cognition) with no explicit and planned coordination create the social, i.e. the emerging of social coordination and the given structures[1]. The dependence on the social structure, surrounding the individuals is not rejected, but plays a minor part in this perspective. As a reaction and critique on the objective (i.e. macro-) perspective and its assumptions of a social organism, of functional adaptation of individuals to the system in the first half of the century, the micro-perspectives received increased attention (see the critiques on Parsons by Schütz, Mead, Blumer etc.). A second motivation was the intention to reduce the scope of society analysis to the social psychology scope of learned behaviour and exchange processes in group theory [21].

Micro-perspective approaches try to investigate how humans typically act under the assumption of the presence of *the generalised other* (see Mead's concept of identity as intersubjectivity and human gestures as *significant symbols)*. These approaches pose the question of which motives and expectations guide the individual's behaviour. They try to reconstruct these motives and expectations from observed situational contexts and behaviour (see the *interpretative paradigm*, a notion which summarises the approaches of Schütz and Mead as well as their followers).

3.3 The Hybrid Approaches

The opposition between micro and macro-approaches belongs to the classical debates of the sociological community. But besides the traditional antonyms corresponding to the micro-macro clash as for example subjectivism-objectivism, system theory vs. theory of action, collectivism vs. individualism etc., we have to note a „renaissance„ of the question about the relation of society (structural aspect) and the individual (action or cognition aspect). The main target of the „hybrid movement„, was to explain social life in relation to both action and the structure, like for example Anthony Giddens did [16].

One of the sociologists with great importance in this respect, not only in France, but all over the world, is Pierre Bourdieu. The conceptualisation of the habitus concept (first 1967) allowed Bourdieu to develop the dialectic relation of objective structure and subjective action/cognition by the assumptions of internalising the structure and reproducing social structure in individual life styles, according to the position in the social space. In contrast to Giddens who created his concept of *structuration structure* for theoretical reasons, the habitus was created and based on Bourdieu's practical work in ethnographic field research in North Africa [7].

---

[1] The following references give an overview to these approaches: The phenomenological approaches in succesion of *Alfred Schütz,* for instance [4] or [42]. For the symbolic interactionism see [32, 5, 18, 13]. For the utilitaristic/behaviourist paradigm see [22, 9, 19].

4 A Brief Summary of the Notion of the Macro-Level in Sociology

Viewing the macro-level of a society means to attribute autonomy to the structural aspects of a social context. These aspects cause stability and change and can be summarised by such concepts as religion, economics, culture, institutions etc. Autonomy here means that no individual does have the power to change these structures and it will even be difficult for a group of individuals. It also means that the structure is not dependent on the existence of a specific individual, the structure survives the individual. While this independence of structure from a specific individual holds, it is also true that the structure depends on the whole population for reproduction of the structure (where reproduction is the only aspect the individuals can influence). It is important to note that this reproduction happens even without explicit knowledge of the individuals. The dynamic that exists in any given social structure is created by the malallocation of resources to individuals. The structures that develop are created as means of reduction of the complexity of life. In this sense society or organisation can only exist if and only if participation of the individual is the „reasonable“ thing to do.

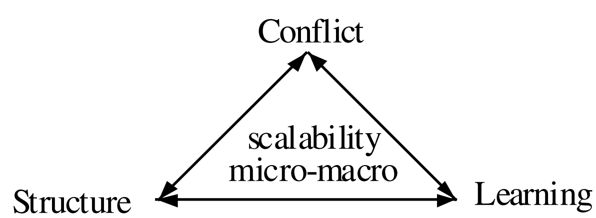


Fig. 2 The relationship of structure, conflict and learning from a *socionics* perspective [31]

In this context it is interesting to remark the connection between structure and learning (or adaptation) on the individual level (see Figure 2). Learning is a cause for structural changes (changing goals, needs and ways of the reproduction of structure) and structure shapes the rules that constrains what and how the individual can learn. This is a connecting point to the idea of the construction of intelligence from the societal context [15]. But there is a second (indirect) connection via the concept of conflict: conflicts are stimuli for learning (e.g. reinforcement learning) and learning may lead to conflicts. Conflict again is connected to structure, as the change of structure often leads to conflicts and conflicts tend to be the causes for such structural changes [27].

5 In Contrast: The Macro Level in DAI

Firstly, we will look at the trends in sociologically motivated agent-based simulation and will give a brief survey of the different applications of the micro-macro distinctions. Secondly, we will look at what can be called application-oriented multi-agent systems. Conte and Moss [12] divide social simulation (not DAI) roughly into these two approaches and we will adopt their terms. The first (sociologically motivated) set of research seeks to develop the foundations of social theory by using DAI in theory testing by simulation, which Conte and Moss call the foundational approach. The other approach, which they name the representational approach, develops modelling techniques and agent specifications to represent observed social and institutional processes. The first set of models and implementations can be viewed as being primarily object to knowledge transfer from DAI to social sciences, whereas the second set may benefit from sociological knowledge in terms of better system performance.

5.1 Agent-Based Social Simulation: The Foundational Approaches

Firstly, there is social simulation research that is inspired by game theoretic approaches, which for instance includes the works that build on Axelrod's research [24],[3]. These works concentrate on modelling attitudes (altruism vs. egoism, benevolence vs. individual rationality) and improve these notions e.g. by mechanisms for protecting cooperative agents from self-interested agents [34]. These works can be viewed as looking at the micro, i.e. interaction level of societies.

A more behaviourist strand of research is the work on platforms like SUGARSCAPE and SWARM (e.g. [23]). Here the macro level is perceived as patterns that emerge from simple behaviours in large sized populations. However, this cannot be attributed as *social actions* as in these models there is no notion of self and others and no action that is intended to influence another individual's belief or actions, which is the very prerequisite of *social* action[2]. A definite exception in this strand of research, are anthropological models that try to elicit emergent structures from social behaviour (behaviour that is directed at other individuals). An example of such research is the EOS project [14], which can demonstrate the emergence of in-group hierarchies, which in sociological terms is a meso-level feature (as the relations of groups are the subject of study).

In the previous approaches the macro-level is perceived as the overall behaviour of a population of agents, an emergent structuring that is not hard-wired by the designer. This is different from the sociological point of view in the respect that sociology would require a number of hierarchies and groups to form, interact and cause changes bi-directionally between micro and macro level.

There is also a strand of research that tries to explicitly model macro structure of a society. However, such multi-level social simulation does not necessarily imply the full bandwidth of sociological concepts of societal levels. For some good reasons (modelling effort, simulation speed) it is common practice to restrict the simulation to only a uni-directional relationship between micro and macro level, which still render impressive results. E.g. Troitzsch [44] describes a multi-level simulation where

---

[2] However, we note that in agent research it is now a common understanding that social ability for an agent does not only mean that the agent can communicate via an agent communication language, but it also implies that the agent is able to model itself and others, reason about *when* to communicate *with whom,* about *what* and in *which way.*

individual (behaviour) was simulated to make predictions about money spending behaviour of a population, attitude formation in a population with no structural changes, gender desegregation in schools etc. In these kinds of simulations the macro level information consists of an aggregation of the micro level data. The design rules out any possibility for the individual to change the structural constraints imposed on the population. According to Conte and Castelfranchi [11] the preference of the uni-directional link for social simulation in current research does not only hold for the micro-to-macro direction but also for the reverse.

5.2 Application-Oriented Multi-Agent Systems: The Representational Approaches

According to Weiß [47], the micro-macro problem poses a question, which raises the issues that define the term of DAI research itself. Therefore we will revisit these issues, before we try to make out important strands of current research and how they relate to the micro-macro discussion. It is important to note that although the micro-macro problem plays such a central role, it is not a standard term in the literature (e.g. [25]). In most of the literature it is referred to only implicitly by trying to decompose the problem into several subproblems.

The first influential collection of such subproblems where we can study at least the implicit notions in DAI of the micro-macro problem is the book by Bond and Gasser [6]. They list five central issues for DAI:

- How to enable agents do decompose their goals and tasks, to allocate sub-goals and sub-tasks to other agents, and to synthesise partial results and solutions.
- How to enable agents to communicate. What communication languages and protocols to use.
- How to enable agents to represent and reason about the actions, plans, and knowledge of other agents in order to appropriately interact with them.
- How to enable agents to represent and reason about the state of their interaction processes. How to enable them to find out whether they have achieved progress in their coordination efforts, and how to enable them to improve the state of their coordination and to act coherently.
- How to enable agents to recognise and reconcile disparate viewpoints and conflicts. How to synthesise views and results.

Please note that compared to the sociological notion of the macro level, these issues are more dealing with agent interaction than societal issues. Moulin and Chaib-Draa [33] add a software engineering (or normative) perspective to this perception of DAI:

- How to engineer and constrain practical multi-agent systems. How to design technology platforms and development methodologies for DAI.

Jennings, Sycara and Wooldridge [25] focus on the coordination aspects in DAI when they add:

- How to effectively balance local computation and communication.

They approach the macro-level from a pragmatic point of view when formulating the last issue for DAI:

- How to avoid or mitigate harmful (e.g., chaotic or oscillatory) overall system behaviour.

This issue is also addressed by a range of game-theory-inspired research, usually summarised under the term *mechanism design* (e.g. [39]). Weiß reformulates these last two issues into the following desiderata:

- How to enable agents to negotiate and contract. What negotiation and contract protocols they use.
- How to formally describe multi-agent systems and the interactions among agents. How to make sure that they are correctly specified.
- How to realise „intelligent" processes such as problem solving, planning decision making, and learning in multi-agent contexts. How to enable agents to collectively carry out such processes in a coherent way.

Especially the last notion seems to be central in DAI: The design of agents that behave coherently. This notion reflects the system designer perspective of a MAS and for the application of MAS we assume that this notion is a cornerstone of the perception of the macro concept. Only occasionally the macro concept is made as explicit in the DAI literature as by Nwana [35]:

> *„macro issues, such as the interaction and communication between agents, the decomposition and distribution of tasks, coordination and cooperation, conflict resolution via negotiation, etc. [The goal of macro research] was to specify, analyse, design and integrate systems comprising of multiple collaborative agents."*

Please note that none of the listed issues deals with the features required by sociology for societies, e.g. power, institutions etc. The term conflict only occurs in the efforts to avoid it (this is the aim of work on *coordination* and *conflict resolution)* and although there is a tremendous concern for patterns of actions, until now there seems to be no theoretical framework to formally analyse such patterns. Verhagen and Smit [45] attribute this to the different approaches of sociology and DAI, where DAI (by its continuingly strong connection to the cognitive sciences) is more concerned about action selection and cognition than the limitations imposed by societal structure on the individual and the effects of knowing about these limitations. Although there is some work on recognising and reasoning about relationships, namely goal/task dependence [40] and role definition and role dependence [20], we cannot say that they approach the far more complex forces that are active on the macro level. Rather, these theories cover the group or organisational level of society.

The confusion of the macro concept between sociology and DAI is partially due to the fact that there is also a (minor) perception of macro as being the structures and rules on the top level of the social context as it is perceived by the individual. In this sense, any given simulated population will have a macro level modelled as well. This holds for prehistoric human communities as well as for even the simplest community model in DAI. However, the majority of social scientists views the most complex level of today's human society as the measure with which the macro level of a population ought to be analysed.

6 Four Misperceptions in DAI Research about Social Phenomena

In this section we apply the sociological notion of the micro and macro level of society for a discussion of the use of these metaphors in DAI research.

1) Mechanism design is macro-level design

Mechanism design is usually the coordination of actions of individuals to achieve some invariants of the behaviour of a group of individuals ([39]; etc.). However, unless there is structure or dynamics in the system that goes beyond the single interaction, there will be no manifestation of societal structures or institutions. In social psychology there is a collection of work inspired by game theory on penalty systems and their emergence in games (e.g. [48]). This could be viewed as advancing to the meso (group) level. Modelling processes among individuals is to be located at the sociological micro level

2) Macro-level behaviour is emergent behaviour

According to Langton [27] emergence is a „result that was not defined statically„ (i.e. before run-time). Such a „not-predefined„ result is not necessarily a macro level result: see for instance SWARM-like simulations. Although they can produce patterns (of action) they do not lead to the emergence of higher-level institutions that shape and keep a society together. A similar argument holds for the reverse direction: macro-level structures can be implemented in a simulation statically without the need to let them emerge.

3) Value aggregation is an analysis of macro phenomena

One way to distinguish attributes for modelling and reasoning, is to differentiate between dimensional (i.e. numerical attributes) and structural (e.g. relationships on cause-effect, or acquaintance, trust, influence etc.). In this differentiation the sociological approach on the macro level (namely to look at structures) is extremely opposed to the one used in current DAI research. The macro perspective here means to aggregate values from the individual to the group layer and focus on dimensional parameters like score, speed, number of communication acts, voting results etc., where aggregation is straightforward. The structural interpretation that could lead to more sophisticated social reasoning, like it is done by Sichmann et al. [33], is rarely applied.

4) Populations of artificial agents are artificial societies

Especially for applied multi-agent systems (the representational approach) it holds that these agents are created with the intention of delegating actions (and in fact *delegation* is viewed as a central notion in DAI: e.g. by Castelfranchi and Falcone [8]). In this sense many assumptions about human behaviour and the user's goals and desires are represented by the agent acting in the multi-agent system. Therefore observed phenomena in this population will not only be caused by artificial actors, but also by the intentions of the human user. As a consequence it would not be correct to speak of an artificial society, the nature of the intersection of intentions requires this to be termed a *hybrid* society. In addition, sociologists would require that this population exhibits macro aspects of the human society (see above) before it can be considered an artificial *or* hybrid society.

7 Towards a Micro-Macro Definition for DAI

We are not in the position to give a final definition for the macro concept for agent-based simulation (either foundational or representational) or decide whether the more complex macro notion of sociology should be applied in DAI. From our research however, we conclude that we can identify three different strands of research where the question of the micro-macro link arises with different magnitude:

a) For moderately sized multi-agent systems (which is still the large majority of today's applications) the list of problem definitions mentioned in Section 4 is sufficiently complex and is most useful due to its well understood distinction in subproblems and its precision.
b) This view is not sufficient when complexity is increased: Open and large multi-agent systems require transfer from the social sciences in order to build systems that are adaptive, scalable and laid out with the potential to resolve unpredictable conflicts. A stronger notion of the macro aspects (institutions, power, fraud etc.) becomes necessary and sociology is a source of inspiration for flexible architectures for scalable MAS. In close analogy to the progress which AI research has made by approaching cognitive psychology, DAI can be expected to be brought forward by the cooperation with sociology.
c) For the knowledge transfer from DAI to the social sciences an adequate conceptualisation of the macro aspect as it is perceived in sociology is necessary to guide agent-based simulation and make the results transferable to sociology.

The approaches in paragraphs a) and b) can be considered representational approaches, whereas c) corresponds to the foundational approach. Paragraph b) views the agent as depending on features like flexibility, autonomy and social competence (where sociologists would argue that the social ability already assumes the flexibility).

Having established that for a number of problems the adoption of a complex and well-founded notion of the macro level is desirable, we would like to discuss some implications for future work.

8 Discussion

A general observation from what has been said so far, is that it may be advisable to use sociologically founded concepts, but computation of bi-directionally interacting micro and macro-level simulation appears to be too complex and too hard to achieve and is therefore hardly existing. When looking at this shortcoming of up-to-date social simulation, it appears that there is a need to investigate, which sociological theory can on the one hand improve the simulated model (e.g. the bi-directional interaction of micro and macro) and on the other hand simplify the design of agents (frameworks for socially more competent agents).

These are the requirements of a hybrid theory that has the explanatory power which stretches from individual behaviour to structures of the social context and back to the individual action. A theory that might come to mind is the theory of Anthony Giddens. The strength of this theory lies in the concept of duality of structure and action. Conte and Castelfranchi [11] criticise that although Giddens' theory „is process-oriented, it actually does not take into sufficient account the role of the cognitive processes linking the micro and macro levels". In our ongoing research we have found that the habitus-field theory of Pierre Bourdieu is a theory which covers a similar spectrum between action and structure, while at the same time having a greater explanatory power on the very subject that Conte and Castelfranchi describe as the shortcoming of Giddens' theory. Bourdieu's concept of habitus consists of a set of dispositions to actions and ways of perception. These dispositions depend on the history of the individual and what it experienced in the past, they may be incorporated or imitated, i.e. learned by observation and acquired by advice. We suggest that the concept of these dispostions is a perfect starting point to connect bounded rationality research with the DAI research of social contexts. Furthermore, the fact that Bourdieu emphasises the practical application of his theory and has reported extensively on his practical work, gives us the hope that his methodology can be used for application in DAI.

For Bourdieu, the habitus is the result of processes that adapt to the surrounding social structure according to the logic of this social context. This marks the importance and the influence of the structure of the agent society on the behaviour of the individual, while still explaining how the individual shapes the structure. Bourdieu views the individual with its desires and actions as the force behind the development, change and reproduction of social structure. For us, this results in a call for more effort in additional reasoning about structures instead of reasoning about aggregated values for agents in social simulation. We believe that with the habitus-field theory we have found a sociological theory that provides what Conte and Castelfranchi [11] demand, when they write:

> *We believe that the micro-macro link is not only a two-fold issue: it is not only a matter of relating macro-structures and micro-interactions, society and action, as many social scientists including Giddens, seem to think. In our conception, it is a three-faceted issue, including (a) external forces and structures, (b) agents' cognition, and (c) their actions. Cognition plays a fundamental linking role between the external forces and the agent's behaviours. ...*
>
> *a) unlike what is commonly called rationality, cognition reflects and embodies in various ways objective pre-conditions, societal prescriptions and institutions, and reinforcing effects. Cognition is undoubtedly structured by society. The question is how is this possible?*
>
> *b) macro-social phenomena may emerge, unintentionally, from micro-interactions. However, they not only directly emerge from behaviours, but also derive from the agent's cognitive representations and state. For example, while some conventions directly emerge from behaviours, some structures of interdependencies emerge from the interrelationships among the internal properties of agents situated in a common world.*

Bourdieu describes his habitus as the structure that is *structured* by the individuals social context and that is also *structuring* the social context by the individuals participation in this context (the „structured and structuring structure"). The incorporation of this structure is the process of learning heuristics for action and perception that are adequate for different contexts. According to Bourdieu these heuristics will not be actively reconsidered before the habitus leads to a crisis. This is an interesting pointer to learning algorithms like reinforcement learning and will guide our future research.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful criticism and ideas on how to improve this document. We would also like to thank the participants of the MABS workshop at the ICMAS 2000 conference for most fruitful discussions. Last but not least, we would like to thank sociologists Michael Florian, Andrea Dederichs and Frank Hillebrandt for their elaborate and constructive comments on an early draft.

References

1. Alexander, J. (1983). Theoretical Logic in Sociology. *Vol. 4:Tthe Modern Reconstruction of Classical Thought.* Berkeley/Los Angeles: University of California Press.
2. Alexander, J. (1995*). Fin-de-siècle Social Theory: Relativism, Reduction and the Problem of Reason,* London.
3. Bazzan, A. L. C., Bordini, R. H. und Campbell, J. A. (1997). *Agents with Moral Sentiments in an Iterated Prisoner's Dilemma Exercise*. In [13].
4. Berger, P. L. and Luckmann, T. (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge.* New York, Doubleday.
5. Blumer, H. (1969). *Symbolic Interactionism.* Englewood Cliffs/New Jersey.
6. Bond, A. H. and Gasser, L. (1988). *Readings in Distributed Artificial Intelligence*, Morgan Kaufmann, pp. 3-35.
7. Bourdieu, P. (1987). *La Distinction: critique sociale du jugement*. Translated by Nice, R.: *Distinction : A Social Critique of the Judgement of Taste*. Harvard Univ Press. 1987
8. Castelfranchi, C. and Falcone, R. (1998). Principles of Trust for MAS: Cognitive Anatomy, Social Importance, and Quantification. In Y. Demazeau. *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS 98)*, 1998.
9. Colemann, J. (1990)*: Foundations of Social Theory,* Cambridge/Mass.
10. Colemann, J., and Bourdieu, P. (1991). *Theory for changing Society*, N.Y.
11. Conte, R. and Castelfranchi, C. (1996). *Simulating Multi-Agent Interdependencies. A Two-Way Approach to the Micro-Macro Link*. In Troitzsch (1996).
12. Conte, R. and Moss, S. (In preparation*). Agent Based Social Simulation: Technological Roadmap*. Document for the AgentLink SIG on Agent-Based Social Simulation.

13. Dautenhahn, K., Masthoff, J. und Numaoka, C. (1997). *Socially Intelligent Agents*. Papers from the 1997 AAAI Fall Symposium, November 8-10, Cambridge, Massachusetts, Technical Report FS-97-02.
14. Doran, J. E. (1997). From Computer Simulation to Artificial Societies. *Transactions SCS*, vol.14(2), pp. 69-77.
15. Edmonds, B. and Dautenhahn, K. (1998). The Contribution of Society to the Construction of Individual Intelligence. In Edmonds, B. and Dautenhahn, K. (eds.), *Socially Situated Intelligence: a workshop held at SAB'98*, August 1998, Zürich. University of Zürich Technical Report, 42-60.
16. Giddens, A. (1984). *The constitution of Society*, Cambridge.
17. Goffmann, E. (1967). *Interaction Ritual*. New York: Anchor Books.
18. Goffmann, E. (1959). *The presentation of Self in Everyday-Life,* N.Y.
19. Goffmann, E. (1961). *Encounters,* Indianapolis, N.Y.
20. Hannoun, M., Sichman, J., Boissier, O. and Sayettat, C. (1998). Dependence Relations Between roles in a Multi-Agent System: Towards the Detection of Inconsistencies in Organization. In Sichman, J., Conte, R. and Gilbert, N. (eds.), *Multi-Agent Systems and Agent-Based Simulation*, LNAI 1534.
21. Homans, C. (1950). *The human group*, New York.
22. Homans, C. (1974). *Social Behavior,* New York.
23. Kennedy, J. (1999). *Artificial Sociocognition: Cultures of Belief in Populations of Elementals*. In Landauer, C. and Bellman, K. L. (eds.): *Virtual Worlds and Simulation Conference (VWSIM'99)*, Simulation Series, vol 31 (2), The Society for Computer Simulation.
24. Jennings, N. R. and Campos, J.R. (1997). Towards a social level characterisation of socially responsibel agents", *IEE proceedings on software engineering* 144(1), pp. 11-25.
25. Jennings, N. R., Sycara, K. and Wooldridge, M. J. (1998). A roadmap of agent research and development. Autonomous Agents and Multi-Agent Systems, 1:7-38.
26. Kuhn, T. (1970). *The Structure of Scientific Revolutions*, Chicago.
27. Langton, C. B. (1989). Artificial Life. *Studies in the Sciences of Compexity,* vol. 6, pp. 1-48, Addison-Wesley.
28. Luhmann, N. (1995). *Social Systems.* Stanford University Press.
29. Luhmann, N. (1997). *Die Gesellschaft der Gesellschaft*. Suhrkamp.
30. Marcenac, P., Courier, R., Calderoni, S. and Soulie, C. (1998). Towards an Emergence Machine for Complex Systems Simulations. In Del Pbil, A. P. and Mira, J. and Ali, M. (eds.): *Lecture Notes in Artificial Intelligence*, vol. 1416, Springer Verlag.
31. Malsch, T. (2000). From a talk given at the kick-off meeting of the German socionics projects in Hamburg Rissen, January 28th.
32. Mead, G. (1934). *Mind, Self and Society,* Chicago.
33. Moulin, B. and Chaib-Draa, B. (1996). An overview of distributed artificial intelligence. In O'Hare, G. M. P. and Jennings, N. R. (eds.): *Foundations of Distributed Artificial Intelligence*, pp. 3-55. John Wiley & Sons Inc., New York.
34. Müller, H.J., Malsch, T. and Schulz-Schaefer, I. (1998). SOCIONICS: Introduction and Potential. Journal of Artificial Societies and Social Simulation. vol. 1(3).

35. Nwana, H. S. (1996). Software Agents: An Overview. In Knowledge Engineering Review, vol. 11 (**3**), pp. 205-244.
36. Parsons, T. and Shills, C. (1962). *Toward a General Theory of Action.* New York.
37. Parsons, T. (1951). *The social system.* Routledge & Kegan Paul, London.
38. Parsons, T. (1971). *The system of modern societies.* Englewood Cliffs, N.J.
39. Sandholm, T. (1999). *Distributed Rational Decision Making*. In Weiss (1999).
40. Sichman, J.S., Conte, R., Castelfranchi, C., and Demazeau, Y. (1994). A social reasoning mechanism based on dependence networks. Proceedings of the 11th European Conference on Artificial Intelligence, Amsterdam.
41. Schillo, M., Rovatsos, M. and Funk, P. (2000). Using trust for Detecting Deceitful Agents in Artificial Societies. In the AAI Journal, Special Issue on "Deception, Fraud and Trust in Agent Societies", eds: Castelfranchi, C., Tan, Y., Falcone, R., and Firozabadi, B. S.
42. Schütz, A. and Luckmann, T. (1973). *The Structures of Life-World.* London: Northwestern University Press.
43. Troitzsch, K., Mueller, U., Gilbert, N. and Doran, J. (1996). *Social Science Microsimulation*. Springer.
44. Troitzsch, K. (1996). *Multilevel Simulation.* In [43].
45. Verhagen, H. and Smit, R. (1997). Multi-agent systems as simulation tools for social theory testing. *Poster presentation at International Conference on Computer Simulation and the Social Sciences (ISSC&SS),* Cortona.
46. Wallerstein, I. (1991). *The Modern World-System: Capitalistic Agriculture and the Origins of the European World-Economy in the sixteenth Century.* Academic Press.
47. Weiß, G. (1999). *Multi-Agent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press.
48. Yamagishi, T. (1986) The Provision of a Sanctioning System as a Public Good. In *Journal of Personality and Social Psychology,* vol. 51 (1), pp. 110-116.

The Simmel Effect: Imitation and Avoidance in Social Hierarchies

Roberto Pedone and Rosaria Conte

IP-CNR, V.le Marx 15, 00137 Roma, Italy
Division "AI, Cognitive and Interaction Modeling" PSS - Project on Social Simulation
{pedone,conte}@ip.rm.cnr.it

Abstract. The simulation study presented in this paper aims to explore the "Simmel effect", i.e. persistence of social differences under instability of status symbols, as an effect of imitation and distinctiveness. A spatial version of the Simmel effect is implemented. Following higher level agents (imitation) combined with moving away from lower level ones (distinctiveness) are found to produce a segregating effect, with agents belonging to the same hierarchical level sharing the same symbols of status.

1 Introduction

Social agents are usually deemed inclined to imitate others. "Attractors", or models of imitation, may represent the majority (cf. majority models), one's closest neighbors (as in Cellular Automata), the most successful or the strongly opinionated agents [1;2]. In multi-agent systems, social laws emerge from agents imitating attractors [3;4].

However, agents not only follow but also distinguish themselves from others. What are the effects of these different social attitudes? An interesting suggestion comes from the work of a famous sociologist, Georg Simmel [5], who described fashion in socially differentiated populations. If agents look after higher level fellows and snub lower-level ones, Simmel hypothesized that status symbols spread through the population downwards, from the highest to the lowest status. As they spread, old symbols are replaced with new ones. Thereby, social differentiation persists under the instability of status symbols. The intuition behind the hypothesis is that agents avoid others to distinguish themselves, and distinctiveness is known to be essential from a biological point of view. The greater the genetic distance among organisms, the larger the variability of the genetic pool and consequently its adaptiveness [6] to a changing environment. By contrast, the effects of distinctiveness are usually disregarded in the social sciences, where agents are viewed as one-drive systems rather than as systems integrating various and possibly antagonistic mechanisms of regulation.

In the next section, our simulation is presented. In the successive section, the hypotheses of the present study, the preliminary experiments conducted, and the findings obtained, are discussed. Ideas for future studies will conclude the paper.

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 149-156, 2000.

2 The Model

2.1 Agents

Agents participating to the simulations are autonomous, they act independently of external intervention. Each agent is modeled as a data type structure with all the information necessary to describe the agent's "body & mind". Agent's data structure embed: agent's body position (x and y coordinate) and agent social hierarchy level.

Agents occupy a "physical" place in the world and can move to others. Two coordinates (Agent.px and Agent.py) identify each agent's current location in the world, represented as a 2-dimension matrix. Agents' positions in the social hierarchy are determined by the value of Agent.g variable. The hierarchy level is represented as an integer number between 1 and 3, with 3 as the highest level and 1 as the lowest.

2.2 Environment

The agents live in a 2-dimension grid of 50x50 cells, representing the world. Each agent occupies a single cell, constantly perceives the nearest agent(s), and can move in four one-step directions, to the north, south, east, and west. More than one agent can stand on the same cell.

2.3 Rules, Strategies, and Simulations

A given level in the social hierarchy characterizes each agent. Each hierarchical level (HL) has associated a different social strategy. Aim of the simulations is to observe whether and under which conditions socially homogeneous clusters of agents emerge. All the simulations involved 250 agents randomly situated in the world, but with the same HL flat distribution (i.e. 84 agents with HL 1, 83 agents with HL 2 and 83 agents with HL 3, cf. fig. 1).

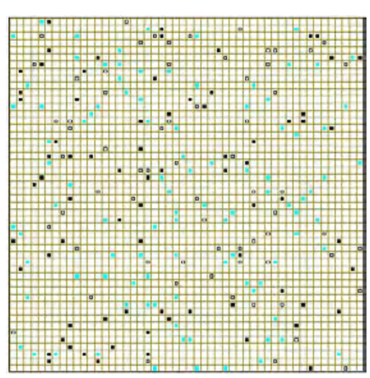


Fig. 1. Simulations' initial configuration

3 Experimental Design

3.1 Hypotheses

The present study reports simulation experiments with a spatial version of the Simmel effect. Agents minimize physical distance from higher level fellows (imitation) and maximize that from lower level ones (avoidance). Simmel suggested that imitation and distinctiveness preserve social differences. Imitation extends higher level symbols to the whole population, thereby reducing differences. On the other hand, distinctiveness acts into the opposite direction: agents differentiate themselves from lower level ones. The two combined rules, therefore, produce the following effects:

- Symbols spread but soon die;
- Hierarchically homogeneous groups emerge (social differences are maintained).

Under the mobility of status symbols, social hierarchy is preserved. Whether this effect is good or bad, is of no concern to us at this stage of our simulation. We aim to explore the dynamics of status symbols and perhaps its cultural-memetic effects [7; 8]. If status symbols are seen as one type of meme, the Simmel effect challenges the view of memes as self-replicating units. Status symbols are bound to be killed as soon as they replicate themselves! A bit theatrically, we could say that their fecundity [7] is their grave [9].

In our simulations, symbols are agents positions on the grid. This is not a perfect metaphor for the Simmel effect. A spatial location is not a symbolic representation and gives rise to neither recombination nor corruption, both usually occurring in the real-world dynamics of status symbols. On the other hand, the spatial metaphor allows the effect to be visualized and to be modeled in its most abstract and simplest form (without corruption of symbols). Moreover, in the real world the spatial segregation often conveys a socially symbolic meaning.

Our model was intended to show the effect of imitation and avoidance on the spatial configuration of social hierarchies. Starting from a random distribution, we expected hierarchically homogenous clusters of agents to emerge thanks to a combined effect of both rules. As the Simmel effect is expected to yield homogeneous groups sharing the same symbols of status, its spatial version was expected to yield spatial vicinity among socially homogeneous agents. In particular,

- With imitation alone, heterogeneous clusters are expected (1st expectation);
- With avoidance alone, no clusters are expected (2nd expectation);
- With both imitation and avoidance hierarchically homogeneous clusters are expected (spatial vicinity of agents sharing the same hierarchical level: 3rd expectation).

3.2 The Experiments

To explore the phenomenon under study, several criteria should be considered:

- Combined vs non-combined rules: both simulations with either imitation or avoidance alone and simulations with mixed rules were run.

- Intersection vs no intersection between followers (imitating agents) and fugitive (escaping agents). Does the Simmel effect applies only when the population is neatly distinguished into a subset of followers and a subset of fugitive with no intersection among them, or does it apply also when there is an intersection between them?
- Social barriers vs. no social barriers: .are social barriers relevant to the emergence of the Simmel effect? Indeed, this effect might be influenced by the existence of social barriers, such that agents can only "perceive" (and then follow/avoid) agents one level up or down in the social hierarchy. And if so, is the Simmel effect conditioned or simply reinforced by social barriers?

Hence, the following simulations result from the crossover of these conditions:

- Imitation alone (simulation 1): the strategy shared by all agents with HL 1 and 2 is to approach (get closer to) other agents with a higher HL. Agents with HL equal to 1 are attracted to HL 2 or HL 3 agents indifferently. HL 3 agents move blindly into a randomly chosen direction. Here, no Simmel effect was expected, but spatial concentration.
- Avoidance alone (simulation 2): the strategy shared by all agents with HL 3 and 2 is move away from agents with lower HL. No Simmel effect was expected, but spatial distribution.
- Imitation + avoidance with no intersection and no social barriers (simulation 3: the main strategy shared by all agents with HL equal to 1 and 2 is approach other agents with higher HL. No social barriers exist. Agents with HL equal to 1 are attracted to HL 2 or HL 3 agents indifferently. The strategy of HL 3 agents is avoid lower HL agents. Here, imitation and avoidance neatly distinguish two subsets of the population, followers (HL 1and 2) and fugitive (HL 3). The Simmel effect was expected to cluster agents 3 as opposed to all others.
- Imitation + avoidance with intersection but no social barriers: (simulation 4): the main strategy shared by all agents with HL equal to 1 is approach any agent with a higher HL (i.e. HL equal to 2 or 3). The strategy of HL 2 agents includes two sub-steps: approach agents with higher HL (i.e. with HL equal to 3) and avoid lower HL agents (i.e. with HL equal to 1). Agents HL 3 avoid any lower HL agents (i.e. with HL equal to 1 or 2). The Simmel effect was expected to form clusters from all the three subgroups (HL 1, 2, and 3).
- Imitation + avoidance with intersection and social barriers: (simulation 5): the main strategy shared by all agents with HL equal to 1 is approach agents with HL equal to 2. The strategy of agents HL 2 includes two steps: avoid agents with HL equal to 1 and approach agents with HL equal to 3. HL 3 agents avoid agents with HL equal to 2. The Simmel effect was expected to form clusters from all the three subgroups and be stronger than in the previous simulations.
- Avoidance + imitation with intersection but no social barriers (simulation 6). Agents with HL equal to 1 approach agents with higher HL (i.e. HL equal to 2 or 3). Avoidance dominates imitation: the strategy of agents HL 2 includes two sub-steps: avoid agents with lower HL (i.e. HL equal to 1) and approach agents

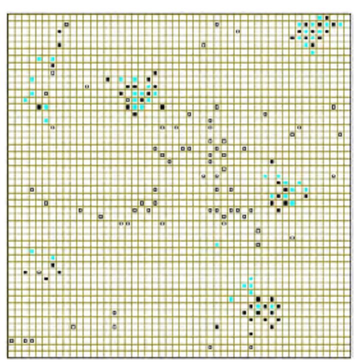
with higher HL (i.e. HL equal to 2). HL 3 agents avoid lower HL agents (i.e. HL equal to 1 or 2). A weak Simmel effect was expected.

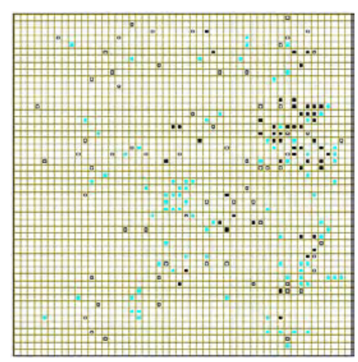
- Avoidance + Imitation with both intersection and social barriers (simulation 7): the main strategy shared by all agents with HL equal to 1 is approach only agents with HL equal to 2. The strategy of agents with HL 2 includes two sub-steps: avoid agents with HL equal to 1 and approach agents with HL equal to 3. HL 3 agents avoid only agents with HL 2.

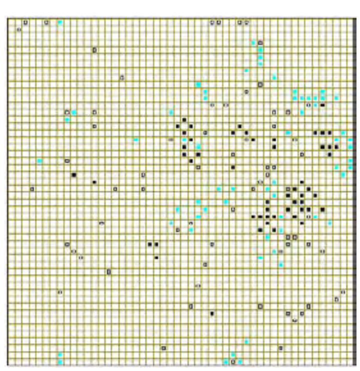
Starting from the same initial random configuration, we have run 5 sets of the 7 different simulations. In each simulation, agents live up to 500 cycles.

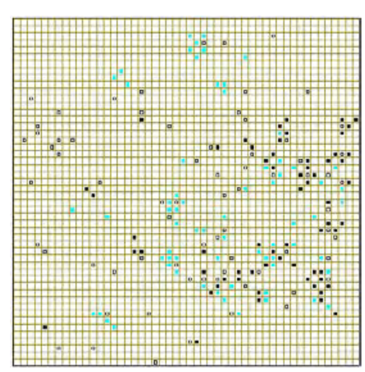
4 Results

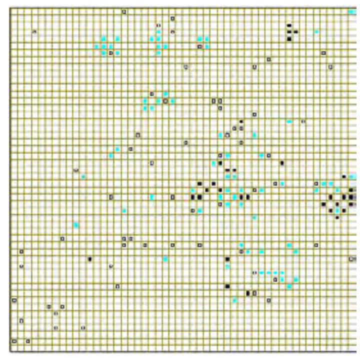
At the end of each simulation, we identified the main visible 'islands' (clusters of aggregated agents) and carried on these a chisquare test to check the hierarchical homogeneity of each of these islands or clusters. This is indicative of how likely hierarchically homogeneous agents tend to clusterize. Results show that in all simulations other than 1 and 2, cluster are significative hierarchically homogeneous ($p<0.05$): cf. Fig. 2.


Simulation 3. Final configuration


Simulation 4. Final configuration


Simulation 5. Final configuration


Simulation 6. Final configuration



Simulation 7. Final configuration

Fig. 2. Simulations 3,4,5,6,7

This confirms our 3rd expectation that the Simmel effect result from both imitation and avoidance. But whilst in simulation 1 (cf. fig. 3) clusters emerge (although non-hierarchically homogeneous), in simulation 2 no clustering effect is visible (cf. fig. 4). This confirms the 1st and 2nd expectations: imitation breaks social differences but yields (spatial) convergence, whilst avoidance breaks social differences and causes spatial dispersion.

Other interesting effects may be drawn from our experiments:

- In the intersection condition, the population will be divided in two clearly distinguished subsets of agents (the followers and the fugitive). No further clustering will appear (cf. Simulation 3 as compared to all other simulations in figure 2).
- Social barriers reinforce the Simmel effect (cf. simulations 5 and 7 as compared to 4 and 6). But they do not condition it: the effect occurs both when agents perceive all others and when they perceive only the nearest in the social hierarchy.

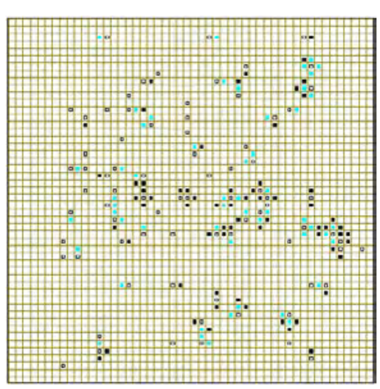


Fig. 3. Simulation 1. Final configuration

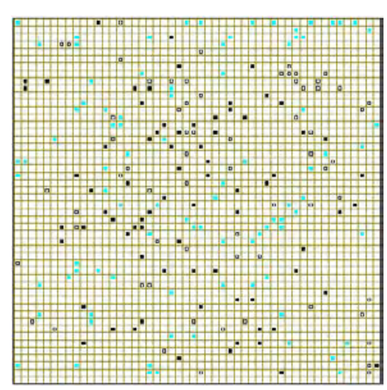


Fig. 4. Simulation 2. Final configuration

- When imitation dominates, the clustering effect is more dramatic (cf. simulations 4 and 5 as compared to 6 and 7 in figure 2) than it is when avoidance dominates. This is consistent with the finding that avoidance has a disruptive effect.

5 Concluding Remarks

In this paper, a simulation model of the effects of imitation and avoidance in social hierarchies has been presented, and some preliminary findings have been discussed. The main objective of the model was to explore a spatial version of the "Simmel effect", i.e. the persistence of social differences and the emergence of hierarchically homogeneous clusters as an effect of the dynamics of status symbols.

The simulation model presented in the paper is based upon a spatial metaphor of the Simmel effect, in which the status symbol is identified with the agents' location on a 2-dimensional grid representing the environment of the simulation. In the simulations described, imitation is implemented as approaching higher level agents, while avoidance is implemented as moving away from lower level ones. In several experimental conditions - with or without intersection between followers and fugitive, with or without social barriers in the hierarchy which limit the agents' social perceptions, and whether imitation dominates avoidance or the other way around - results confirmed our expectations. The Simmel effect takes place when both rules co-exist. Furthermore, the effect is reinforced by but is not conditioned to social barriers. With intersection, clusters get formed from all social levels rather than from the highest level only. Finally, the clustering is more visible when imitation dominates avoidance than with the opposite pattern, since avoidance has a desegregating effect.

These findings encourage the simulation study of the Simmel effect. First, this study helps to highlight the minimal conditions for social segregation. Second, it questions the oversimplified view, held by social as well as multi-agent scientists, that social agents are essentially ruled by imitation, and calls for a more complex view where social agents are ruled by imitation and distinctiveness. Third, it points to the interplay between social and cultural processes, which may interact without necessarily presenting the same qualitative effects. Indeed, cultural dynamics may co-exist and even favor the maintenance of the social status quo. Fourth, the simulation study of the Simmel effect has a memetic side-effect: it addresses issues concerning the transmissibility of memes. In the phenomenon under study the fecundity of memes seems to ultimately lead to their extinction, a finding which is in apparent contrast with the memeticists' thesis that memes are self-replicating units.

In future studies, a symbolic, rather than spatial, representation of status tokens ought to be adopted in order to examine the effects of corruption and recombination of status symbols, which seems a rather unavoidable aspect of the phenomenon under study.

References

1. Latané, B. (1981) The psychology of social impact. American Psychologist, 36:343-356.
2. Rockloff, M.J. & Latan□, B. (1996) Simulating the social context of human choice. In K. G. Troitzsch, U. Mueller, N. Gilbert, J. Doran (eds) Social Science Microsimulation, Berlin: Springer.
3. Shoham, Y. & Tenneholtz, M. (1992). Emergent conventions in multiagent systems: Initial experimental results and observations. Proc. of the 3rd Intern. Conf. on KR&R. Cambridge, MA, 225-232.
4. Walker, A. & Wooldridge, M. (1995). Understanding the emergence of conventions in multi-agent systems, Proc. of the 1st Intern. Conf. on Multi-Agent Systems, MIT Press, 384-389.
5. Simmel, G. The Fashion (first published in 1895 in Germany).
6. Cavalli Sforza, L.L. & Feldman, M. (1981) Cultural Transmission and Evolution. A Quantitative Approach. Princeton, N.J: Princeton Univ. Press.
7. Dawkins, R. (1976) The Selfish Gene. Oxford Univ. Press.
8. Blackmore, S. (1999) The Meme Machine. Oxford: OUP.
9. Benvenuto, S. (2000) Dicerie e pettegolezzi. Bologna: Il Mulino.

Cooperation without Memory or Space: Tags, Groups and the Prisoner's Dilemma

David Hales

Department of Computer Science, University of Essex
Colchester, Essex, UK
daphal@essex.ac.uk

Abstract. A recent [14] model demonstrated that image scoring produces high cooperation between strangers in the Prisoner's Dilemma (PD). Here we outline a simpler approach in which players – which are either pure cooperators or defectors – can sustain cooperation with strangers by biasing game interactions towards others with the same tags (arbitrary bit strings representing "cultural markers" [11]). In our model there is no requirement for knowledge of past performance or recognition of individual players. Unlike spatial models [13] reproduction of strategies is population wide. Contrary to previous tag models [15] cooperation is demonstrated in the single round game.

1 Introduction

In modern and complex social worlds, individuals are required to interact cooperatively with many strangers using limited knowledge and bounded rationality. But why do strangers cooperate? Here we discount those situations in which cooperation is possible without trust and examine that subset of cooperative interactions that follow the form of a social dilemma [10]. We formalise this kind of encounter using the ubiquitous form of the two player single round Prisoner's Dilemma game. We demonstrate empirically, for the first time, that the biasing of game interaction towards agents sharing identical tags (arbitrary markings represented as bit strings) is sufficient to produce high levels of cooperation in the single round PD when agents are boundedly rational optimisers. Interestingly, this process can be visualised as the formation and dissolution of "groups" that share the same tags in a non-physically extended abstract "tag space". Firstly we introduce the Prisoner's Dilemma game, then we outline some existing theories which attempt to explain the emergence of cooperation within such a game between evolutionary optimisers. We then introduce our minimal tag based model (the TagWorld) and give the results obtained. Finally we discuss the significance of the results and their possible applicability to human societies.

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 157-166, 2000.

2 The Prisoner's Dilemma

The Prisoner's Dilemma (PD) game models a common social dilemma in which two players interact by selecting one of two choices: Either to "cooperate" (C) or "defect" (D). From the four possible outcomes of the game payoffs are distributed to the individuals. A reward payoff (R) and a punishment payoff (P) are given for mutual cooperation and mutual defection respectively. However, when individuals select different moves, differential payoffs of temptation (T) and sucker (S) are awarded to the defector and the cooperator respectively. Assuming that neither player can know in advance which move the other will make and wishes the maximise her own payoff, the dilemma is evident in the ranking of payoffs: $T > R > P > S$ and the constraint that $2R > T + S$.

Although both players would prefer T, only one can attain it. No player wants S. No matter what the other player does, by selecting a D move a player ensures she gets either a better or equal payoff to her partner. In this sense a D move can't be bettered since playing D ensures that the defector can not be suckered. This is the so-called "Nash" [8] equilibrium for the single round game. It is also an evolutionary stable strategy [12] for a population of randomly paired individuals playing the game where reproduction fitness is based on payoff. But the dilemma remains, if both individuals selected a cooperative move they would both be better off. But many societies (human and animal) appear to have solved (at least some) dilemmas similar to the PD. How can this be explained by purely evolutionary mechanisms?

3 Evolutionary Extensions

Evolutionary selection favours selfish individual replicators. When collections of these replicators form groups it's possible for them to co-ordinate their behaviour in ways which would make global optimisation possible. The kinds of behaviours that make this possible include, cooperation, altruism and specialisation. All of these are observed in animal and human societies. But evolutionary selection does not seem to offer an explanation for these behaviours. To address this problem three extensions of natural selection have been proposed: kin selection [9], group selection [17] and reciprocal cooperation [2]. Although each offers explanations of some of the kinds of the social behaviours of interest neither seems to offer a general framework applicable to human or artificial social systems. Kin selection only applies to highly genetically related individuals, group selection in its simplest form is fundamentally flawed (selfish individuals within the group benefit relative to altruists) and reciprocal cooperation does not explain true altruism (i.e. cooperative behaviour in the single round PD). Neither does it scale-up well to large groups due to the cognitive demands from the requirement that all interactions be on-going with recognisable individuals and associated memory of past interactions.

4 The TagWorld

Agents are represented as fixed length bit strings (of length L+1) comprising a tag of length L bits and a single strategy bit. The strategy bit represents a pure strategy, either unconditional cooperation or unconditional defection. Initially the population of agents are set to random bit strings (with each bit decided by a fair coin toss). The following evolutionary algorithm is then applied:

```
LOOP some number of generations
  LOOP for each agent (a) in the population
    Select a game partner agent (b)
    Agent (a) and (b) invoke their strategies receiving
      the appropriate payoff
  END LOOP
  Reproduce agents in proportion to their average payoff
END LOOP.
```

In each generation each agent (a) is selected from the population (of size N) in turn. A game partner is then selected. Partner selection entails the random selection of another agent (b) from the population such that (a) ≠ (b) followed by a comparison of tags with agent (a). If the tags are identical a game interaction takes place otherwise (b) is returned to the population without game interaction. If (b) was returned to the population without interaction a second (b) is selected at random from the population and its tag compared with (a). This process is repeated until an agent (b) is found which has an identical tag to (a) or an upper limit F of selections has been reached. If this upper limit is reached then game interaction is forced on the next randomly chosen agent. Consequently (a) will always find a partner even if its tag does not match any other agent because an agent which can not find a matching partner will eventually exhaust its upper limit F of refusals and then be forced to interact with a randomly chosen partner.

During game interaction (a) and (b) invoke their strategies and receive the appropriate payoff. After all agents have been selected in turn and played a game a new population is asexually reproduced. Reproductive success is proportional to average payoff. The entire population of N agents is replaced using a "roulette wheel" selection method [5]. Equation 1 and inequality 2 outline this method. Equation 1 gives the total average payoff for the entire population where a_i is the *ith* agent from the population, *ap(a)* is the average payoff obtained by agent *a*, and N is the size of the population. The inequality 2 specifies an agent a_x to select for reproduction from the population where *x* is the smallest integer that satisfies the inequality and *rnd(0..tap)* is a uniformly randomly selected value in the range *0..tap*. The inequality is satisfied N times with a different random value. Each time, *x* gives the index of an agent to reproduce. Using this method the probability that an individual will be reproduced into the next generation is proportional to average payoff.

$$tap = \sum_{i=1}^{N} ap(a_i) \tag{1}$$

$$rnd(0..tap) \leq \sum_{i=1}^{x} ap(a_i) \tag{2}$$

Mutation is applied to each reproduced player with probability M = 0.001. This low value indicates the assumption that agents rarely change their strategy. Since there are 100 agents in the population, we would expect one strategy bit to change over 10 generations (on average). Mutation takes the form of flipping each bit of each player with probability M. Consequently *tags and strategies* are mutated in reproduced agents.

The PD payoffs are parameterised over T (the temptation payoff for defectors over cooperators) such that T > 1. The reward R for mutual cooperation is 1. The punishment P for mutual defection and the sucker payoff S for cooperation with a defector are both 0.0001. This value was selected because it was small but greater than zero (indicating a very small chance for agents, with Sucker or Punishment payoffs, of reproduction). If a small value is added to P (enforcing T > R > P > S) results are not significantly changed.

5 Results – High Cooperation

A set of runs to 100,000 generations with a population of size N = 100 agents was executed for various values of T and L. The maximum number of refusals was fixed at F = 1000. This high value of F means that it is unlikely that agents will not be paired with other matching agents (if they exist) in the population. For the purposes of analysis cooperation was characterised as the proportion of mutually cooperative interactions occurring over all generations. This figure was calculated by counting the number of games in which both agents cooperated for the whole run (of 100,000 generations) and then dividing by the total number of games played. Thus the level of cooperation for a single run is derived from the results of 10^7 individual games.

Figure 1 shows results for various values of L and T graphically. Each bar represents an average of 5 independent runs. As the values of L are increased, and T are decreased, cooperation increases monotonically. As can be seen, where L >= 32 very high levels of cooperation are obtained for all values of T.

The results obtained indicate that very high levels of cooperation can be sustained between optimising agents in the single round PD via simple tag biasing. There is no requirement for knowledge of past performance or recognition of individual agents (i.e. the other agents may be viewed as strangers). Unlike some spatial models which have demonstrated cooperation in the single round game [13] reproduction is population wide. The next section explains this high cooperation in terms of group formation.

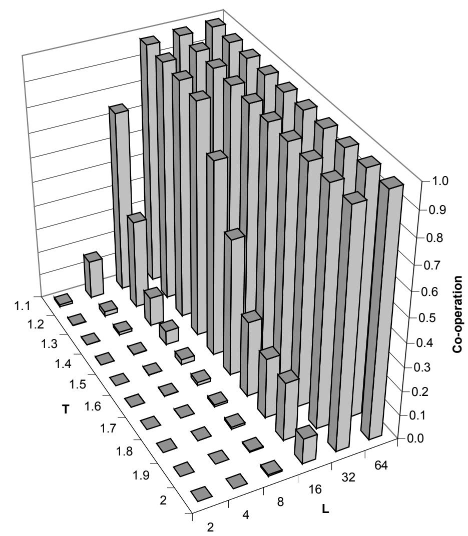


Fig. 1. Cooperation for various values of L (number of tag bits) and T (the temptation payoff) are shown. Cooperation is measured as the proportion of mutually cooperative games over 100,000 generations. Each point is an average of 5 runs. The entire chart represents 3×10^9 individual game interactions

6 Group Formation and Dissolution

The tag space can be visualised as an L-dimensional hyper-cube with corners representing unique tag values. Agents sharing a tag, share a corner. Mutation produces movement between corners. Game interaction is therefore taking place in an abstract "tag space". Cooperative groups sharing matching tags will form in corners of the hyper-cube. These groups will outperform non-cooperative groups and hence tend to increase in size over generations. However, if mutation introduces defecting agents into a cooperative group they will tend to outperform the cooperators within the group (by suckering them). From this the seeds of the destruction of the group are planted, since as the number of defectors increases within a group the overall fitness of agents within the group decreases. Other more cooperative groups (if they exist) will tend to expand. While this process is occurring, mutation of tag bits will produce a slow migration of agents between corners of the hyper-cube, possibly founding new groups in previously empty corners.

Figure 2 is a visualisation of the process over time taken from a single run. Each line on the vertical axis represents a corner of the hyper-cube (i.e. unique tag value). The horizontal axis represents time in generations. If no agents have a particular tag value in a given generation then the line is left blank (white). Alternatively, if a corner contains all cooperative agents then the line is light grey. For a mixed group in which there are both cooperators and defectors the line is dark grey. For an entirely defective group the line is black. Examination of figure 2 shows the time evolution of groups in tag space. Initially cooperative groups (light grey lines) become invaded by defectors producing mixed groups (darker grey) which very swiftly become entirely defective (black) and then quickly go extinct (white).

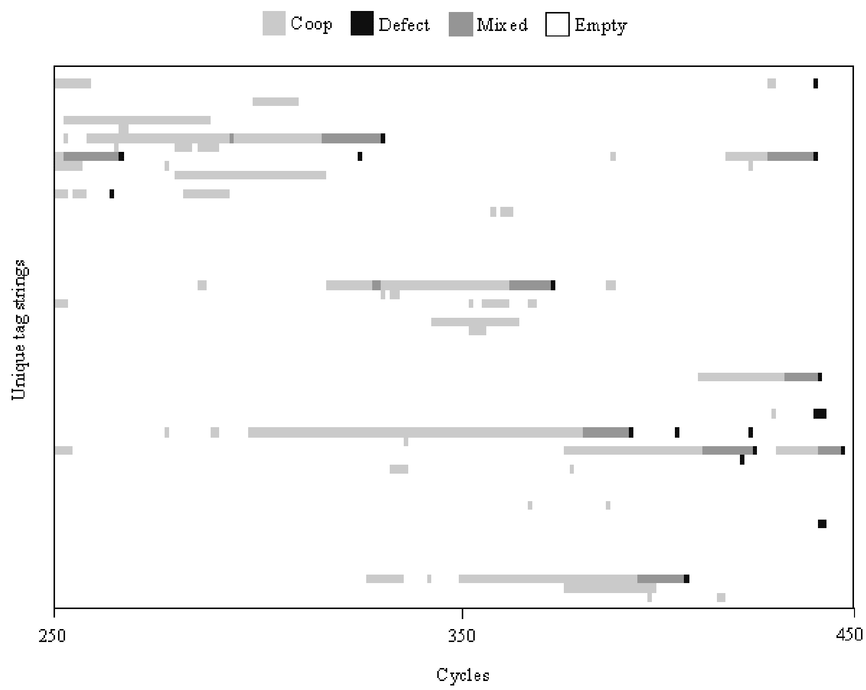


Fig. 2. Visualisation of 200 cycles (generations) from a single simulation run showing cooperative groups coming into and going out of existence. Each line on the vertical axis represents a unique tag value (of which only a subset is shown). If all agents sharing a tag value are cooperative then the line is light grey. If all agents are defectors then the line is coloured black. A mixed group is shown as dark grey. The horizontal axis represents time in cycles (generations). Here L = 8 and T = 1.1

7 Cooperation from Complete Defection

Although high levels of cooperation are demonstrated over many contiguous generations, starting from the random initialisation of agents, these results do not indicate if a society can recover cooperation from a state of complete defection. In order to test this, experiments were conducted in which the initialisation of agents was modified so that all agent strategy bits were set to defection. Figure 3 shows a set of runs for various values of N (population size) against the number of generations before mutual cooperation emerges. These empirical results are compared to a simplified analytical (probability based) treatment given in equation 3.

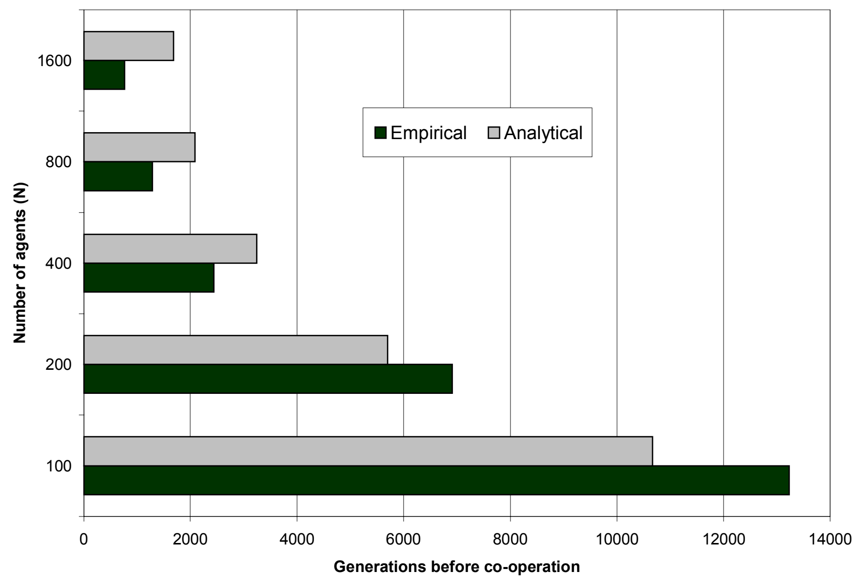


Fig. 3. Number of generations before mutual cooperation emerges. Here a comparison is made between the analytical model (given in equation 3) and the empirical results. Each bar for the empirical test comprises an average of 10 runs. Tag size L = 32 for all runs shown

$$ang(n,m) = \frac{1}{\frac{2}{n-1}(1-(1-m)^n - nm(1-m)^{n-1})} \tag{3}$$

Equation 3 gives the expected average number of generations required (from an initial society of all defectors) before two (and only two) cooperative agents perform a game interaction. This is dependent on the population size N and the mutation rate M. In all of the 50 runs used to form the empirical results in figure 3 it was observed that high cooperation immediately (in the next generation) followed the first occurrence of a mutually cooperative encounter between two agents. It was also observed that drift over the tag bits tended to lead a society of all defectors toward sharing the same tag bits. These empirically observed (rather than derived) phenomena were used to simplify the analytical treatment: we assume all agents share the same tag bits and high cooperation starts when two agents mutate to cooperators and game interact.

In figure 3 the predictions of equation 3 are compared empirically with actual runs. As can be seen the analytical treatment tends to under-estimate the number of generations before cooperation emerges for low population sizes and over-estimate for large population sizes. It can be hypothesised that this is due to the simplification in equation 3. Drift will rarely produce populations in which all agents have identical tags. When the population is split between different tag values the simplification will under-estimate the number of generations required. This is because two cooperators may be produced but they may not share the same tag and hence will not game interact. Conversely when the population size is large it will often be the case that more than two agents will be mutated to cooperators. A more detailed analysis would need to capture the dynamics of the tag bits over the population and the effect of those dynamics on producing cooperative interactions.

8 Discussion

In a previous study of tag based partner selection in the PD, Riolo [15] concluded that tags produced little increase in cooperation in the single round game. In his work a tag was represented as a single real number attached to each agent. The abstract topology of the tag space was therefore one dimensional. The matching of tags was based on a probabilistic function applied to the distance between two tags, meaning that agents with similar but not identical tags could engage in game interactions. The number of refusals allowed before forced interaction was low (50 refusals in a population of 200 agents). Additionally a fitness cost was attached to each refusal made by an agent (although this was reduced to zero in an attempt to get high cooperation in the single round game). Under these conditions it was demonstrated that high cooperation emerged when agents engaged in the Iterated PD (IPD) but not in the single round PD game. The interpretation placed on this previous study was of agents representing animals, searching for game partners and evolving genetically. In the work presented here allowable refusals is high (10 times the population size) and there is no associated cost. Also the tag is represented as a bit string which must match *exactly* with a partner for game interaction to be selected by an agent. Under these assumptions a different kind of tag space topology is possible and high cooperation is produced in the single round game.

Intuitively it would seem that the *exact tag matching* constraint is not necessary to produce high cooperation in all cases. For example, consider a situation in which mutation was zero and two groups of agents existed in the tag space such that there was no inter-group game interaction. If one group consists of all cooperators and the other contains some defectors then the cooperative group would expand at the expense of the non-cooperative group. This would even hold if agents were applying some partial matching scheme - *so long as there is an interaction boundary between the two groups*. By "interaction boundary" is meant that some mechanism partitions the agents into strict game interaction groups. That games can only take place between individuals sharing a group.

The interpretation placed on this work is of agents representing a "bundle" or "complex" of culturally learned and transmissible behaviours, a so-called "meme-complex" [3], [7]. The assumption is that in a population of hosts meme-complexes which produced high utility (for their hosts) would be more likely to be copied (in proportion to their relative utilities). One key to such a process producing high altruism and co-operation is the packaging of the tag with the beneficent strategy or behaviour as a cultural unit. Allison in his theory of altruism [1] echoes this assumption with reference to the importance of "cultural packaging techniques". In the work presented here "packages" of tags and defective strategies do not dominate the population because such packages destroy the very groups that they are a part of.

In recent work Bowles & Gintis [4] give a detailed analytical treatment of the value of groups in the promotion of co-operation in the PD when binary social cues convey useful predictive information concerning a game partners strategy. However, they do not address the issue of *how* social cues come to be have such predictive utility. The results produced within the TagWorld society show that even simple mechanisms can produce this kind of correlation because groups which contain defectors quickly die out. In contrast to the examples used by Bowles & Gintis (who focus on racial groups) the mechanism which produces this quick extinction of non-cooperative groups within the TagWorld society requires that cultural interaction, in the form of individuals moving between groups easily, is high. It is this ability of individuals to quickly swap cultural groups, by taking on new tags from others, which drives the co-operation producing process. Strong group boundaries which prevent easy entry and exit from a group would hinder or even destroy co-operation forming by the process illustrated in the TagWorld.

The TagWorld was parameterised over a number of dimensions. A scan was made over a restricted part of the parameter space demonstrating that high co-operation is present over a that area of that space. However, other dimensions of the space have not been explored. The role of mutation and refusals would be an interesting area of investigation. Also, measures other than just co-operation, such as group sizes over time and migration rates between groups, would be of interest and could be used to elaborate an analytical model. This could link this work with patch based models of altruism developed within evolutionary biology [17]. Another interesting area of further investigation would be to make refusals an endogenous parameter encoded into each agent and able to evolve. In such a scenario would high refusal rates evolve? If so then the assumption of high refusals could be justified via endogenous evolutionary processes rather than as an exogenous assumption of the model.

Acknowledgements

This work would not have been possible without the help of, discussions with, and suggestions from, Prof. Jim Doran, Department of Computer Science, University of Essex, Colchester, UK. I am also grateful to Rick Riolo (Program for the Study of Complex Systems, The University of Michigan, Ann Arbor, MI, USA) for his perceptive and helpful comments and suggestions on a earlier draft of this paper. This work was supported by an EPSRC research studentship award and is based on work for a forthcoming PhD thesis.

References

1. Allison, P.: The cultural evolution of beneficent norms. Social Forces, **71**(2), (1992) 279-301.
2. Axelrod, R.: The Evolution of Cooperation. Basic Books, New York (1980)
3. Blackmore, S.: The Meme Machine. Oxford University Press, Oxford (1999)
4. Bowles, S. & Gintis, H.: Optimal Parochialism: The Dynamics of Trust and Exclusion in Networks. Santa Fe Institute Working Paper 00-03-017. Santa Fe, NM (2000)
5. Davis, L.:Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York (1991)
6. Dawkins, R.: The Selfish Gene. Oxford University Press, Oxford (1976)
7. Dennett, D.: Darwin's Dangerous Idea. Simon & Schuster, New York (1995)
8. Gibbons, R.: A Primer In Game Theory, Harvester, New York (1992)
9. Hamilton, W.: Altrusim and Related Phenomena, Mainly in Social Insects. In: Annual Review of Ecology and Systemics **3** (1972) 193-232
10. Hardin, G.: The tradigy of the commons. In: Science. 162 (1968) 1243-1248
11. Holland, J.: The Effect of Lables (Tags) on Social Interactions. Santa Fe Institute Working Paper 93-10-064. Santa Fe, NM (1993)
12. Maynard Smith, J.: Evolution and the Theory of Games. Cambridge University Press, Cambridge (1982)
13. Nowak, M. & May, R.: Evolutionary Games and Spatial Chaos. Nature, **359**, (1992) 532-554
14. Nowak, M. & Sigmund, K..: Evolution of indirect reciprocity by image scoring. Nature, **393**, (1998) 573-557.
15. Riolo, R.: The Effects of Tag-Mediated Selection of Partners in Evolving Populations Playing the Iterated Prisoner's Dilemma. Santa Fe Institute Working Paper 97-02-016. Santa Fe, NM (1997)
16. Simon, H. A.: A Mechanism for Social Selection and Successful Altruism. Science, **250** (1990) 1665-1668
17. Wilson, E. O.: Sociobiology: The New Synthesis. Harvard University Press, Cambridge Mass (1975)

Sexual Attraction and Inter-sexual Dominance among Virtual Agents

Charlotte K. Hemelrijk

Department of Information Technology and Anthropological Institute and Museum,
University of Zürich, Switzerland
hemelrij@ifi.unizh.ch

Abstract. In many group-living primates, males are dominant over females, but despite this dominance, they allow females access to resources during the period when females are sexually attractive - but only then and not otherwise. Conventionally, such male 'courtesy' is explained as a special strategy to gain mating access to females. In the present paper I propose a simpler hypothesis that is based on an agent-centered model, namely that male 'courtesy' to females is in fact a kind of 'timidity' that arises because sexual attraction automatically increases female dominance. The model consists in a homogeneous, virtual world with agents that group and perform dominance interactions. VirtualMales have a higher intensity of aggression and start with a greater capacity to win conflicts than VirtualFemales. I shall explain how the addition of attraction of VirtualMales by VirtualFemales leads to female dominance, and other phenomena that are relevant to the study of animal behaviour.

1 Introduction

In many animal species, males are extremely attracted to females, whereas females are relatively uninterested in males [1]. Male guppies are a good example [2]. They spend almost all their time courting and only if there is direct danger of a predator, they may pause. Although primate males are less ardent, males are the ones who actively maintain proximity to females when females are in their sexually attractive, oestrus period [e.g. see 3]. This sexual asymmetry is understandable, because males can fertilise many females, whereas females get fertilised only once per reproductive period. To obtain access to females, males have been supposed to develop many strategies. For instance, primate males are observed to allow oestrus (but not anoestrus) females priority of access to food sources [4-6]. This is regarded as an intentional manipulation by the male [7] and as an adaptive exchange of favours, namely priority of access to food for females in exchange for copulation for males. Evidence for such exchange is, however, very limited, if existing at all [8] and probably not every behavioural act should be interpreted separately in terms of costs and benefits to the number of offspring [9]; Surely, simpler alternative explanations are badly needed [10].

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 167-180, 2000.

Here, artificial life models come in as a useful tool, because they show that animal behaviour is determined not only by specific inherited properties and individual intentions, but also emerges by self-organisation from simple behavioural rules that lead to very complex behaviour in unexpected ways from the feedback between qualities of individuals and their changing environment. Whereas in the behaviour of real organisms, the effects of different contributing factors cannot be distinguished, they can be more easily traced in an Artificial World, for several reasons. An Artificial World can be used as a kind of Virtual Laboratory in which behavioural rules and parameters can be changed at will and the consequences analysed. Further, the behavioural rules of the artificial agents and the changes in their variable features are fully known. This makes such a model a suitable tool for developing our knowledge of the rules that may underlie observed behavioural patterns.

The aim of the present paper is to study whether in the absence of benefits and of calculative intentions, female dominance and, consequently, male 'tolerance', increases by self-organisation more strongly when males are attracted to females than when they are not. Besides, I investigate some other patterns of social behaviour usually studied by ethologists.

This paper is one of a series of papers [e.g. 11, 12, 13], in which I have shown that unexpected behavioural patterns, such as cooperation and exchange [14, 15] emerge by self-organisation in a model of competing, group-living artificial agents. Such a model represents only a few important features of an animal society [and is originally inspired by 16]. It consists of a homogeneous artificial world inhabited by agents that are equipped merely with a tendency to group and to perform competitive interactions. The effects of winning and losing such interactions are self-reinforcing [as has also been empirically observed in many animals species, for references see 10, 17]. To make the effects of experienced winning and losing of the agents and of pure chance as clear as possible, all agents are at the start completely identical. Yet, a dominance hierarchy develops over time. This development appears to depend on the intensity of aggression and the cohesion of grouping. For medium and higher values for both [18], a steep hierarchy develops in mutual reinforcement with a spatial structure with dominants in the centre and subordinates at the periphery [12]. This feedback has many unexpected consequences that all resemble observed behaviour of despotic species of primates, e. g. macaques. However, when the hierarchy is weak, the society resembles that of egalitarian macaques [19]. This correspondence makes the model a suitable tool for generating hypotheses for real primates.

In my model, I have introduced artificial 'sexes', by creating two types of agents that differ exclusively in their competitive ability [20]. In line with descriptions of primates [21], aggression by artificial males is made more intense than that by artificial females (implying more frequent biting compared to slaps and threats). Furthermore, reflecting the physiologically superior fighting ability of males (e.g. muscle structure), artificial males start with a higher tendency to win than artificial females. Unexpectedly, in a society with a steep hierarchy artificial females appear to dominate more males than in a society with a weak hierarchy. This arises as a side-effect of the stronger differentiation of dominance values for each sex separately [19]. Up till now, however, my models have ignored, the fact that in real animals, during certain periods of the year females are sexually attractive and males are highly interested in them.

In the present model, I study the consequences of this attraction. Hereto, I introduce attraction of VirtualMales to VirtualFemales and examine its effect on inter-sexual dominance relations and other variables studied previously [such as cohesion, spatial centrality of dominants, hierarchical differentiation and frequency of aggression e.g. see Hemelrijk 14, 18, 20]. I shall show how indeed dominance of artificial females inevitably increases and how this can be used as a parsimonious alternative for the evolutionary hypothesis of sexual exchange. At the same time, females become more aggressive and, as another side effect, artificial males appear to develop more aggression among themselves.

2 Methods

In this section, a description of the model and behavioural measures is given.

2.1 The Model

The model is individual-oriented and event-driven [see 22]. I have written it in object-Pascal, Borland Pascal 7.0 and it consists of three parts: a 'world' with its interacting agents, its visualisation and special observers that collect and analyse data on what happens in the 'world' (cf. the 'recorders' and 'reporters' of Hogeweg, 1988). The 'world' has the form of a torus (a seemingly three-dimensional donut) to avoid border effects and consists of a space of 200 by 200 units. At the start of each run agents occupy random locations within a predefined subspace of 30 by 30 units. The space of the world is made continuous, in the sense that agents are able to move in all directions. This continuous world is used because it represents spatial patterns more precisely than a grid world, which I applied formerly [32]. Agents have an angle of vision of 120 degrees and their maximum perception distance (MaxView) is 50 units. Parallel simulations cannot be run on most computers and therefore, activities of agents are regulated by a timing regime. Studies have shown that a specific timing regime influences the results of the simulation [33]. Often a random regime is applied in which each entity receives a random waiting time from a uniform distribution and the one with the shortest waiting time is activated first. Here, I combine a random regime with a biologically plausible timing regime that is locally controlled by other entities [see also, 34,35]. The locally controlled timing regime reduces the waiting time of an agent if a dominance interaction occurs within the agent's NearView (24 units). A nearby dominance interaction is thus considered as a kind of 'disturbance' that increases the chance that the agent is activated. This reflects observations on real animals, whereby dominance interactions are likely to activate individuals nearby [compare social facilitation, see 36]. Agents group and perform dominance interactions according to a set of rules described below (Figure 1).

Grouping Rules. Usually, two opposing tendencies affecting group structure are supposed to exist: on the one hand animals are believed to be attracted to one another because participation in a group provides safety; on the other, aggregation implies competition for resources, and this drives individuals apart [e.g., 23].

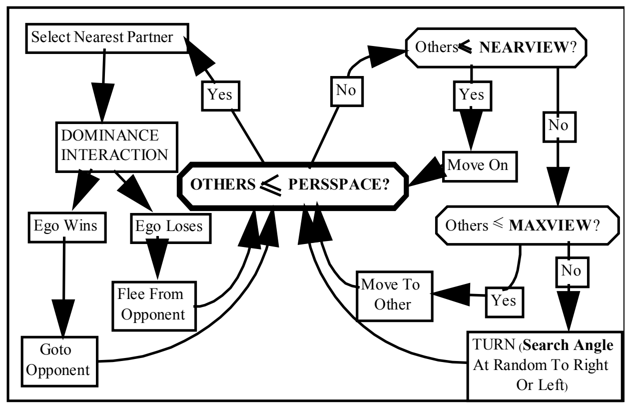


Fig. 1. Flow chart for the behavioural rules of agents that are not attracted to another type (sex)

The tendency of individuals to aggregate and space out are represented in the model by a set of rules that are graphically displayed in Figure 1 [see 19].

1. If an agent observes another within a critical distance, its 'personal space' (= PerSpace, see 2.2), it may perform a dominance interaction. If several agents are within PerSpace, the nearest interaction partner is chosen. If the agent wins the interaction, it moves one unit towards its opponent, otherwise it makes a 180° turn and flees away two units under a small random angle.
2. If nobody is observed in PerSpace, but an agent notices others at a greater distance, but still within NearView (see 2.2), then in runs without 'sexual attraction', it continues moving one unit in its original direction. In case of 'attraction', however, VirtualMales approach a VirtualFemale one unit when they observe her in nearView.
3. If its nearest neighbours are outside NearView, but within its maximum range of vision (= MaxView, see 2.2), the agent moves towards one unit them.
4. If an agent does not perceive other agents within MaxView, it looks around for them by turning a Search angle of 90° at random to the right or left.

Dominance Interactions. Dominance interactions in real animals consist of competitive interactions about nearby resources (such as food, mates and spatial location), but seem not always motivated by competition for immediate resources and some interactions are considered part of a kind of long-term 'power' struggle. In the model, these two types of dominance interactions are not distinguished and resources are unspecified. Dominance interactions may be initiated when agents encounter each other nearby, but happen only if the perceived risk of defeat is low [in the so-called risk sensitive behaviour, 19]. Interactions between agents are modelled after Hogeweg [24] and Hemelrijk [19], as follows:

Each agent has a variable that is called 'Dom' (= dominance, representing the capacity to win an interaction).

After meeting one another in their PerSpace, agents 'decide' whether or not to attack following the Risk-Sensitive system. Here, the probability to attack decreases according to the risk of defeat as follows. Upon meeting another agent and observing its Dom-value, an agent may foresee it will win or lose on the basis of a 'mental' battle, which follows the rules of a dominance interaction as described below. If ego loses the mental interaction, it will refrain from action (thus displaying 'non-aggressive' proximity). If it wins the mental battle, it will start a 'real' dominance interaction.

If an actual dominance interaction takes place, then agents display and observe each other's Dom. Subsequent winning and losing is determined by chance and values of Dom as follows :

$$w_i = \begin{cases} 1 & \dfrac{DOM_i}{DOM_i + DOM_j} > RND(0,1) \\ 0 & \text{else} \end{cases} \tag{1}$$

Here w_i is the outcome of a dominance interaction initiated by agent i (1=winning, 0=losing). In other words, if the relative dominance value of the interacting agents is greater than a random number (drawn from a uniform distribution), then agent i wins, else it loses. Thus, the probability of winning is greater for whoever is higher in rank, and this is proportional to the Dom-value relative to that of its partner.

Updating of the dominance values is done by increasing the dominance value of the winner and decreasing that of the loser:

$$\begin{aligned} DOM_i &:= DOM_i + \left(w_i - \frac{DOM_i}{DOM_i + DOM_j} \right) * STEPDOM \\ DOM_j &:= DOM_j - \left(w_i - \frac{DOM_i}{DOM_i + DOM_j} \right) * STEPDOM \end{aligned} \tag{2}$$

The consequence of this system is that it functions as a damped positive feedback: a victory of the higher ranking agent reinforces its relative Dom-value only slightly, whereas success of the lower ranking agent gives rise to a relatively great change in Dom. The impact thus reflects the degree to which the result is unexpected. (To keep Dom-values positive, their minimum value is, arbitrarily, put at 0.01.) The change in Dom-values is multiplied by a scaling or stepping factor, so-called StepDom, which varies between 0 and 1 and represents intensity of aggression. High values imply a great change in Dom-value when updating it, and thus indicate that single interactions may strongly influence the future outcome of conflicts. Conversely, low STEPDom-values represent low impact. This study is confined to high values near 1.

Winning includes chasing the opponent over one unit distance and then turning randomly 45 degrees to right or left in order to reduce the chance of repeated interactions between the same opponents. The loser responds by fleeing under a small random angle over a predefined FleeingDistance.

In what follows, the initiation of a dominance interaction is for short referred to as 'attack'.

2.2 Experimental Set-Up and Data Collection

Here, the same parameter setting (n=8, persSpace=2, nearView=24, SearchAngle 90, FleeingDistance=2 units) is used as in a former study [18].

The present study is confined to a population size of eight agents consisting of two types that differ in fighting capacity. VirtualMales start with a higher winning tendency than VirtualFemales (i.e. of 16 versus 8) and display a higher intensity of aggression (i.e. StepDom value of 1.0 versus 0.8).

Two conditions (with and without attraction to females, see 2.1.1.'Grouping Rules') are compared. In the condition of 'sexual attraction' all females are supposed to be attractive, whereas in the condition without attraction none of them is. For both conditions 10 runs are conducted, resulting in a total of 20 runs.

During a run, every change in spatial position and in heading direction of each agent is recorded. After every time unit (consisting of 160 incidences of activation), the distance between agents is measured. Dominance interactions are continuously monitored by recording (1) the identity of the attacker and its opponent, (2) the winner/loser and (3) the updated Dom-values of the agents.

2.3 Measurements

At intervals of two time units (320 incidences of activation), the degree of rank differentiation and the overlap between the dominance hierarchies of VirtualMales and VirtualFemales are measured as follows.

Dominance differentiation is measured by the coefficient of variation (standard deviation divided by the mean) of Dom-values [25]. For each run the average value is calculated. Higher values indicate larger rank distances among agents.

At the start of each run, all VirtualMales are dominant over all VirtualFemales, but during run-time some VirtualFemales may become dominant over (some or all) VirtualMales. The degree of dominance of VirtualFemales over VirtualMales is estimated by the Mann Whitney U- statistic [26]. Hereto, for each female the number of males ranking below her are counted. The value of the statistic is calculated as the sum of these countings. At the beginning of the run U-values are zero. Later on they may become positive.

The clustering together of agents of the same sex is measured as a τ_{Kr}-correlation between a matrix of mean distance among agents and a 'hypothesis'-matrix. The 'hypothesis'-matrix reflects sexual-segregation because cells belonging to agents of the same sex are filled with the number 1 and cells of different sexes are filled with zeros. Segregation is thus reflected by a positive correlation.

The spatial direction in which others are located as regards a certain individual ('ego') is used as a measure of the degree with which individuals occupy the centre. Using circular statistics [27] the centrality of each individual is calculated for each scan by drawing a unit circle around it and projecting the direction of other group members (as seen by ego) as points on the circumference of this circle. Connecting these points with the origin produces vectors. The length of the mean vector represents the degree in which the position of group members relative to ego is clumped; longer mean vectors reflect more clustering in one direction and indicate lower centrality (i.e. lower 'encirclement'). Thus, greater centrality of higher-ranking agents is reflected in a stronger negative correlation between rank and encirclement.

Differences in behaviour between societies with and without attraction to females are tested using one data-point per run, namely the mean frequency of interaction per time-unit per sex. To exclude a possible bias brought about by transient values, the correlation for centrality of dominants is calculated on data collected after time-unit 200.

3 Results

3.1 Effects on Female Dominance

When VirtualMales are attracted to VirtualFemales in NearView, this increases dominance of VirtualFemales over VirtualMales markedly, as shown by the larger number of males being subordinate to VirtualFemales compared to that when 'attraction' is absent (Figure 2A). Simultaneously, the mean dominance value of VirtualFemales is higher and of VirtualMales is lower than in the absence of attraction to females (Figure 2BC). Consequently, the mean dominance values of both sexes converge during attraction. In its absence, however, they diverge over time (Figure 2D).

Previous Explanations of Female Dominance. How does female dominance over males arise? In earlier models, cohesion (via a steeper hierarchy) and the frequency of interaction between the sexes have been shown to contribute to female dominance [18-20].

Cohesion contributes to female dominance via the accompanying hierarchy, which is steeper than in loose groups [18], as follows. Hierarchical differentiation develops together with spatial centrality of dominants. It appears to arise from the positive feedback between the development of the hierarchy and spatial centrality of dominants. A clearer hierarchy implies that the lowest ranking agents flee from about everyone else. Therefore, they end up at the periphery leaving the dominants in the centre. When agents are thus sorted according to dominance, they will usually meet and interact with partners of close rank. This implies that if a dominance reversal occurs at all, it will only be a minor one. Thus, spatial sorting according to rank stabilises the hierarchy and keeps the differentiation intact. This feedback (clearly leading to bi-directional causation, see Sawyer, same volume) develops more clearly under the spatial constraints and high frequency of interaction of cohesive than of loose grouping. Strong differentiation of dominance values will automatically cause some of these females to obtain a higher dominance position than some males, even though artificial females start with a weaker tendency to win than males. Conversely, weak hierarchical differentiation will leave artificial females in their initial subordinate position to males.

Besides, female dominance also appears to arise in the presence of a weaker hierarchy, if the sexes interact more often with each other than among themselves [as reflected by a weaker correlation for segregation of the sexes, see 18].

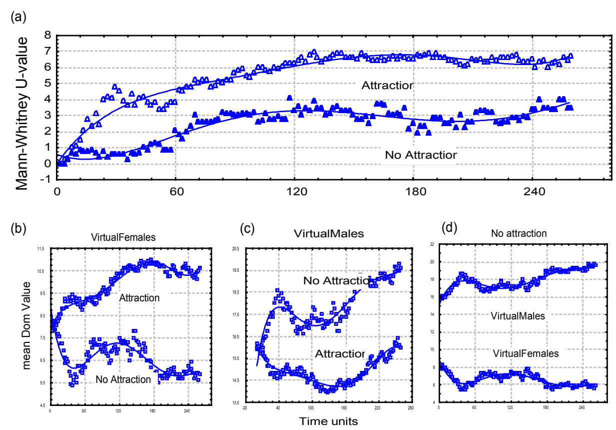


Fig. 2. Dominance differentiation for the two societies with and without attraction of VirtualMales to VirtualFemales. (a) Dominance of VirtualFemales over VirtualMales measured by the Mann Whitney U-statistic as the summed countings of the number of males ranking below each female. (b),(c),(d). Mean dominance values calculated over 10 runs for individuals of the same sex

In summary, in the cases described above, greater female dominance arises from stronger cohesion which goes hand in hand with more marked spatial centrality of dominants and stronger hierarchical differentiation and from weaker sexual segregation.

None of these two processes can, however, explain the present finding of increased female dominance due to male attraction to females, for the following reasons.

First, whether VirtualMales are attracted to females or not, the group cohesion (Mann Whitney U-test, $n_a = n_n = 10$, U=46, P=0.762), the hierarchical differentiation (Mann Whitney U-test, $n_a = n_n = 10$, U=32, P=0.174) and the spatial structure (Mann Whitney U-test, $n_a = n_n = 10$, U=28.5, P=0.103) remain similar. Thus, there is no indication of a stronger social-spatial feedback. Second, adding attraction between the sexes, does unexpectedly, not increase the relative frequency with which the sexes interact with each other and among themselves, as measured by the correlation for sexual segregation (Mann Whitney U-test, $n_a = n_n = 10$, U=46.5, P=0.796). Obviously, it is a different process that causes female dominance in the present case and therefore, it is necessary to look in greater detail at patterns of aggression.

Alternative Explanation for Female Dominance. Despite the unchanged cohesion, attraction to VirtualFemales raises the frequency of attack in the group (Mann Whitney U-test, attack: $n_a = n_n = 10$, U=0, P=0.00016, Figure 3a). This fact explains the curious finding that attraction does not increase cohesion. Apparently for the present parameters, stronger cohesion is counter-balanced by the increased frequency with which agents drive each other apart. Note, further, that in contrast to previous results [18] the higher frequency of aggression does not increase spatial structure and differentiation of the hierarchy. The cause of this may be that the movement by VirtualMales counteracts sorting by rank, because they approach any VirtualFemale independently of her dominance.

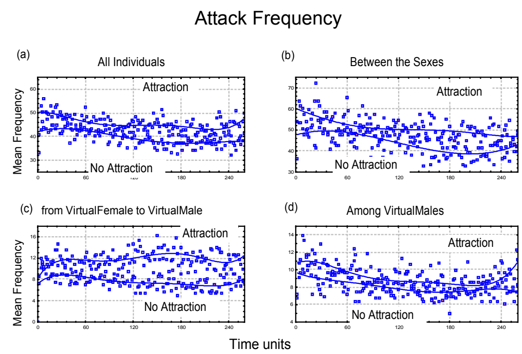


Fig. 3. Mean Attack frequency of the two societies with and without attraction of VirtualMales to VirtualFemales. Mean dominance values calculated for 10 runs for all individuals of the same sex

More specifically, aggression is increased between the sexes both in its absolute frequency and its percentage of total aggression (Mann Whitney U-test, frequency: $n_a = n_n = 10$, U=0, P=0.00016, Figure 3b; percentage: $n_a = n_n = 10$, U=21, P=0.029). It is this that triggers female dominance as an implication of the rule inbuilt in the model, that the degree with which the outcome of a fight changes the dominance values of both partners, depends on the degree to which the outcome of the conflict was expected. Dominance values of both partners undergo a greater change if, unexpectedly, a lower-ranking agent defeats a higher-ranking one than if, expectedly, a subordinate is beaten by a dominant. As a consequence, defeat of dominants by subordinates produces dominance conversion of the 'sexes' more strongly than expected victories by dominants induce divergence of dominance of both sexes. Since the higher percentage of interaction between the sexes implies a higher percentage of incidental victories of VirtualFemales over VirtualMales at the beginning, attraction will accelerate dominance conversion between both types (the 'sexes').

Greater dominance of VirtualFemales over VirtualMales makes VirtualFemales more aggressive, particularly against VirtualMales (Mann Whitney U-test, $n_a = n_n = 10$, U=19, P=0.019, Figure 3c), than if VirtualMales are not attracted to them. Note, that this may further enhance their dominance (Kendall Rank Correlation between mean female aggression and dominance, n=10, Tau=0.733, P=0.003).

3.2 Other Consequences

When VirtualMales are attracted to VirtualFemales this also raises the frequency of non-aggressive approach in the group ($n_a = n_n = 10$, U=8, P=0.0015, Figure 4a). Due to their lessened dominance VirtualMales more often approach others (Mann Whitney U-test, $n_a = n_n = 10$, U=4, P=0.0005, Figure 4b) and particularly VirtualFemales non-aggressively (Mann Whitney U-test, $n_a = n_n = 10$, U=12, P=0.0041, Figure 4c).

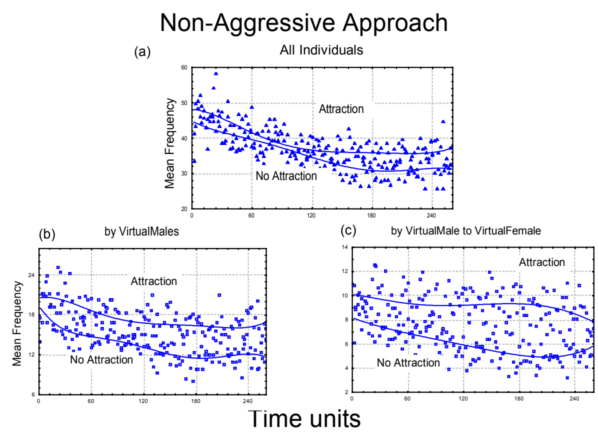


Fig. 4. Mean frequency of non-aggressive approach of the two societies with and without attraction of VirtualMales to VirtualFemales. Mean dominance values calculated for all individuals of the same sex and for 10 runs

When VirtualMales are attracted to VirtualFemales, this raises also intra-sexual aggression among VirtualMales (Mann Whitney U-test, $n_a = n_n = 10$, U=4, P=0.0001, Figure 3d). Although this needs to further study, it probably comes about via an increased opportunity to meet each other close by when several VirtualMales converge on the same VirtualFemale. Consequently, VirtualMales will more often trespass on each other's personal space and therefore, attack. In line with this, there is also a non-significant trend that VirtualMales more often approach each other non-aggressively (Mann Whitney U-test, $n_a = n_n = 10$, U=32, P=0.19).

Note that the increased frequency of intrasexual aggression explains why, as mentioned above, the correlation for sexual segregation of interactions remain similar with and without attraction: whereas inter-sexual aggression increases as expected, this is not reflected in a weaker correlation, because it is neutralised by the increase in aggression frequency among VirtualMales.

4 Discussion

When in this society of group-living artificial agents, VirtualMales are attracted to VirtualFemales, there are several unexpected consequences.

First, such attraction increases the dominance of VirtualFemales and decreases that of VirtualMales. This is due to the higher frequency of interaction between the sexes and the inbuild mechanism that unexpected victories and defeats cause a greater change in the dominance values of both opponents than expected outcomes do. This mechanism is based on precise behavioural observations of dominance interactions in bumblebees [28] and seems a plausible assumption for any species. It provides a new mechanism by which female dominance over males increases. Note that this process (together with a steeper hierarchy) will also contribute to female dominance in cohesive groups, because compared to loose groups, the frequency of intersexual interactions and thus also of incidental victories by the weaker sex, will be higher.

Second, VirtualMales behave more often non-aggressively to VirtualFemales. The greater dominance of VirtualFemales makes it more risky for VirtualMales to attack them. Thus, the model presents us with a parsimonious alternative for the usual explanation of the observed male 'tolerance' towards females at food sites when females are in their receptive period and males are highly attracted to them. The conventional adaptive explanation for this is, that males increase their number of offspring by exchanging food for sex [5, 6], but the model shows how even in the absence of any benefits accrued to 'tolerant' male behaviour, attraction to VirtualFemales may produce male 'tolerance' (or rather timidity) via the increase of female dominance over males.

Third, VirtualFemales become more aggressive when VirtualMales are attracted to them. Similarly, primate females are described as being more aggressive when in oestrus [e.g. see 6, 29]. Whereas this may be due to their special hormonal state as is traditionally supposed, the model suggests two alternative, more simple mechanisms that may be operative: an increase of encounter frequency with males and, consequently, increased female dominance over males.

Fourth, and unexpectedly, attraction to VirtualFemales makes VirtualMales more aggressive among themselves though they do not want to monopolise VirtualFemales! Although this will be studied in further detail in the near future, it is probably a consequence of the increased frequency of their meeting near VirtualFemales: VirtualMales will more often enter each other's attack range and actually attack. Similarly, in a combined modelling and empirical study of butterflies [Euphydryas anica, 30], male grouping and increased male aggression are suggested to result as a side-effect of male mate-searching behaviour. Male butterflies typically investigate anything that even remotely resembles a female. Such indiscriminate searching causes males to investigate each other. Resulting male-male chases cause a change in the direction of their movement, which, particularly under high density, lead to male aggregations and increased aggression.

Obviously, the model does not represent the complexity and sophistication of real animals. It does not even represent social positive and sexual behaviour. Instead, the model just incorporates the self-reinforcing effects of dominance interactions among agents that are grouping indiscriminately apart from the fact that males move preferentially towards females during certain periods. This simplification is useful, because the model represents features that are relevant for many animal species and it makes the implications of these assumptions detectable, which cannot be done in studies on real animals due to the many unknown variables. Note that the condition of sexual attraction in the present model concerns attraction to all females, suggesting that all of them are synchronously tumescent. Although this holds for some primate species, in others, female menstrual cycles are not synchronised. The effects of such asynchrony will be studied in a model in future.

Further, I am at present studying the same model for the case where females are attracted to males, whereas males are indifferent. Although this situation hardly appplies to any animals in the real world, these models can also represent 'species' that do not exist [31] and allow us to search the 'world of the possible' and thereby, reveal unknown general processes. From the present paper and my former ones it appears that one cannot predict how these changes will affect social structure, in terms of variables such as interaction frequencies, spatial configuration and dominance overlap between the sexes.

In summary, in this and former models, I have presented evidence that inter-sexual dominance may be influenced by the intensity of aggression, by cohesion and by sexual attraction. At first sight, however, a connection of these three variables with intersexual dominance is far from obvious. It seems hardly possible to arrive at this kind of explanations by decomposing behaviour in independent components as is usually done. This shows that individual-based models are indispensable tools to obtain hypotheses how social behaviour, whether in animals or in humans, may emerge from the feedback between changing features of individuals and their group members by self-organisation.

Acknowledgements

I am grateful to Rolf Pfeifer and Bob Martin for continuous support. I like to thank Jaap Hemelrijk for correcting the English. This work is supported by the Swiss National Science Foundation by a grant from the Marie Heim-Vögtlin Foundation and by a grant from the A. H. Schultz foundation.

References

1. Trivers, R.L., *Parental investment and sexual selection*, in *Sexual selection and the descent of man*, B. Campbell, Editor. 1972, Aldine: Chicago. p. 136-179.
2. Magurran, A. and M. Nowak, *Another battle of the sexes.* Proceedings of the Royal Society London, 1991. 246: p. 31-38.
3. Hill, D., Social relationships between adult male and female rhesus macaques: 1. Sexual consortships. Primates, 1987. 28: p. 439-456.
4. Yerkes, R.M., *Social dominance and sexual status in chimpanzees.* Quarterly Review of Biology, 1939. 14: p. 115-136.
5. Yerkes, R.M., Social behavior of chimpanzees: dominance between mates in relation to sexual status. Journal of Comparative Psychology, 1940. 30: p. 147-186.
6. Goodall, J., *The chimpanzees of Gombe: patterns of behaviour*. 1986, Cambridge MA and London: Belknapp Press of Harvard University Press.
7. de Waal, F.B.M., *Foodsharing and reciprocal obligations among chimpanzees.* J. Human Evolution, 1989. 18: p. 433-459.
8. Hemelrijk, C.K., C.M. Meier, and R.D. Martin, *'Friendship' for fitness in chimpanzees?* Animal Behaviour, 1999. 58: p. 1223-1229.
9. Gould, S.J. and R.C. Lewontin, *The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptionist programme.* Proceedings of the Royal Society London B, 1979. 205: p. 581-598.
10. Hemelrijk, C.K., Towards the integration of social dominance and spatial structure. Animal Behaviour, 2000, 59, 1035-1048.

11. Hemelrijk, C.K., *Reciprocation in apes: from complex cognition to self-structuring.*, in *Great Ape Societies*, W.C. McGrew, L.F. Marchant, and T. Nishida, Editors. 1996, Cambridge University Press: Cambridge. p. 185-195.
12. Hemelrijk, C.K., *Spatial centrality of dominants without positional preference*, in *Artificial Life VI*, C. Adami, *et al.*, Editors. 1998, MIT-Press: Los Angeles, USA. p. 307-315.
13. Hemelrijk, C. Risk Sensitive and Ambiguity Reducing Dominance Interactions in a Virtual Laboratory. in Proceedings of the Fourth International Conference on Simulation on Adaptive Behavior. From Animals to Animats 5. 1998. Zürich: MIT Press.
14. Hemelrijk, C.K., Dominance interactions, spatial dynamics and emergent reciprocity in a virtual world, in Proceedings of the fourth international conference on simulation of adaptive behavior, P. Maes, et al., Editors. 1996, The MIT Press: Cambridge, MA. p. 545-552.
15. Hemelrijk, C.K., *Cooperation without genes, games or cognition.*, in *4th European Conference on Artificial Life*, P. Husbands and I. Harvey, Editors. 1997, MIT-Press: Cambridge MA. p. 511-520.
16. Hogeweg, P., *MIRROR beyond MIRROR, Puddles of LIFE*, in *Artificial life, SFI studies in the sciences of complexity*, C. Langton, Editor. 1988, Adisson-Wesley Publishing Company: Redwood City, California. p. 297-316.
17. Bonabeau, E., G. Theraulaz, and J.-L. Deneubourg, *Mathematical models of self-organizing hierarchies in animal societies.* Bulletin of Mathematical Biology, 1996. 58(4): p. 661-717.
18. Hemelrijk, C.K., Effects of cohesiveness on intersexual dominance relationships and spatial structure among group-living virtual entities, in Advances in Artificial Life. Fifth European Conference on Artificial Life, D. Floreano, Nicoud, J-D., Mondada, F., Editor. 1999, Springer Verlag: Berlin. p. 524-534.
19. Hemelrijk, C.K., *An individual-oriented model on the emergence of despotic and egalitarian societies.* Proceedings of the Royal Society London B: Biological Sciences., 1999. 266: p. 361-369.
20. Hemelrijk, C.K., Self-reinforcing dominance interactions between virtual males and females. Hypothesis generation for primate studies. Adaptive Behavior, in press.
21. Bernstein, I.S. and C.L. Ehardt, *Intragroup agonistic behavior in Rhesus monkeys (Macaca mulatta).* International Journal of Primatology, 1985. 6(3): p. 209-226.
22. Judson, O.P., *The rise of the individual-based model in ecology.* Trends in Ecology and Evolution, 1994. 9: p. 9-14.
23. Schaik, C.P.v., *Why are diurnal primates living in groups?* Behaviour, 1983. 87: p. 120-144.
24. Hogeweg, P. and B. Hesper, *The ontogeny of interaction structure in bumble bee colonies: a MIRROR model.* Behavioral Ecology and Sociobiology, 1983. 12: p. 271-283.
25. Sokal, R.R. and F.J. Rohlf, Biometry: the principles and practice of statistics in biological research. 2 ed. 1981, San Francisco: W.H. Freeman.
26. Siegel, S. and N.J. Castellan, *Nonparametric statistics for the behavioral sciences*. second ed. Statistics Series. 1988, New York: McGraw-Hill international editions. 399.
27. Mardia, K.V., *Statistics of directional data*. 1972, London: Academic Press.

28. Honk, C.v. and P. Hogeweg, *The ontogeny of the social structure in a captive Bombus terestris colony.* Behavioral Ecology and Sociobiology, 1981. 9: p. 111-119.
29. Michael, R.P. and D. Zumpe, Aggression and gonadal hormones in captive rhesus monkeys (Macaca mulatta). Animal Behaviour, 1970. 18: p. 1-10.
30. Odendaal, F.J., P. Turchin, and F.R. Stermitz, *An incidental-effect hypothesis explaining aggregation of males in a population of Euphydryas anicia.* The American Naturalist, 1988. 132(5): p. 735-749.
31. Ray, T.S., An approach to the synthesis of life., in Artficial Life II, Santa Fe Institute, Studies in the sciences of compexity, C.G. Langton, et al., Editors. 1991, Addison-Wesley, CA: Redwood City. p. 371-408.
32. Hemelrijk, C. K. 1997 Cooperation without genes, games or cognition. In *4th European Conference on Artificial Life* (ed. P. Husbands & I. Harvey), pp. 511-520. Cambridge MA: MIT-Press.
33. Huberman, B.A. and N.S. Glance, *Evolutionary games and computer simulations.* Proceedings of the National Academy of Sciences of the United States of America, 1993. 90: p. 7716-7718.
34. Goss, S. and J.L. Deneubourg, Autocatalysis as a source of synchronised rhythmical activity in social insects. Insectes Sociaux, 1988. 35(3): p. 310-315.
35. Hogeweg, P. and B. Hesper, *A model study on biomorphological description.* Pattern Recognition, 1974. 6: p. 165-179.
36. Galef, B.G.J., Imitation in animals: history, definitions, and interpretation of data from the psychological laboratory., in Social learning: Psychobiological and biological perspectives, T. Zentall and B. Galef, Editors. 1988, Erlbaum: Hillsdale, New Jersey. p. 3-28.

Multi-Agent Modelling and Renewable Resources Issues: The Relevance of Shared Representations for Interacting Agents

Juliette Rouchier[1], Francois Bousquet[2], Olivier Barreteau[3], Christophe Le Page[2], and Jean-Luc Bonnefoy[4]

[1] Centre for Policy Modelling - Manchester Metropolitan University, UK
j.rouchier@mmu.ac.uk
http://www.cpm.mmu.ac.uk/~juliette/
[2] Cirad-Tera-Ere
Montpellier, France
bousquet,lepage @cirad.fr
[3] Cemagref
Montpellier, France
barreteau@montpellier.cemagref.fr
[3]Université de Provence, France
Jean-Luc.Bonnefoy@up.univ-aix.fr

Abstract. The issue that is addressed in this paper concerns the way interacting agents should understand their environment so that a common good used by the whole group would last. We synthesise the results of four models with agents interacting in artificial societies in which they have to share a resource. The four societies were built using multi-agent based simulation models that address issues related to the use of common renewable goods. The resources that are used by the artificial communities of agents are of two types: for some, agents must co-ordinate to exploit the resources; for others, the distribution of goods among agents is directly dependent on the distribution of the agents in space. But that classification cannot necessarily hold: the good use of the resources relies on an even distribution of agents in space, but this can be obtained with individual processes in some cases whereas in others it implies coordination too.

1 Introduction

1.1 Issue

The issue that is addressed in this paper concerns the way interacting agents should understand their environment so that a common good used by the whole group would last. This topic is related to fundamental issues in the two domains that we were addressing in our research: multi-agent systems and renewable resources economy. The "representations" we refer to in the paper are linked mainly to the definition of

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 181-197, 2000.

the Agents' research field: it is the understanding that the agent has of what it perceives, and that enables it to choose its actions.

The paper is based on results of simulations, made in multi-agent artificial societies [3], [10], [18]. In all the models the environment is a common good that the agents have to use to survive. Because the models were realised by different people, they not only describe different case studies but they also exhibit different theoretical approaches to the issue. They are all gathered here because the building of specific representations for agents in relation to an evolving environment was their common, main question. In all of them there was a serious concern that one should model individual representations of the environment so that the agents display some expected global behaviours. The most important issue was then to know how much the agents should share in these representations. A "representation" is here the value that an agent has of an action it could undertake on its environment in a distributed way: should it use or not some elements of its environment. What interests us here especially is the idea that the agents have of what is attracting for them: should they all agree, or have very different approaches to the same object or agent?

1.2 Theory and Tools

A usual approach to the "best way" to use a resource in the main stream of ecological economics, is to consider that the necessary exchanges of goods are going to lead to an equilibrium that imply a good use of the resource. In that case, what is important is that the agents produce different goods and are interested in getting different ones, too, and thus that they are willing to exchange [8].

In a multi agent system, agents can have their own representations of the world, evolving independently of the representations of the others, and deduced from their interactions with other agents and the actions of those other agents. Alternatively, it is possible to impose some shared representation, where the agents always behave according to norms and common habits. Actually, the majority of the research in the field of MAS for collective actions deals with joint beliefs, joint intentions and joint commitment [12] In order to solve complex tasks in common, the underlying assumption is that agents must share a common representation of the problem even though they have different roles [7].

1.3 Applications

In our applications, the representations are the information that the agents use to choose their actions concerning the common resource and the other agents. In all cases, the agents have an individual representation of their environment and there can exist a common representation. The individual representation of the agents can be close to or different from the common one; it can evolve in an endogenous way or be stable in the model.

As a first analysis, the different models give results that could seem to be quite contradictory: in three models the agents must have different evaluations of the same resource so that the global dynamics is sustainable; in the other, only common points of view can insure a good use of the environment. That is the reason why we had to look for a more global explanation. The assumption was that it was the characteristics of the resource that would explain that difference.

The models describe several kinds of goods and resources:

- In Shadoc the agents use an irrigated system to water their field. They must first secure their access to credit and then their access to water to keep cultivating from one time-step to another.
- Djemiong, where agents have to hunt in a forest. They can gather in groups or decide to act individually. The sustainability of the use of the resource is then studied depending on the way they all act.
- JuMel describes the use of grazing lands by mobile herders performing exchanges of goods with farmers so that to get access to pastures and water. All exchanges are made among individuals through negotiations.
- CommonForest describes a community where each agent has her cattle graze in a forest. There is a common decision that spots the places to protect, and the agents have the choice to obey to that common prescription, or to behave like they individually want to.

In Shadoc, the agents share a common production tool and thus require a common representation to use this tool in common for a long time. For two models where resource is distributed, we show that the agents must have heterogeneous representations if the resource is to be used in a sustainable way. In the last model the agents have an homogenous point of view of what their actions must be, but the individual evaluation of the resource (the representation they have of it) differs a lot among them. This indicates that the best way to model the agents' representation of the environment is related to the way the production itself is represented, and this choice is of course directly related to the type of resources that are studied.

2 Models

In this section we present all four models, describing the context of their building, what an agent is in each, what kind of representation it has of its surroundings and of the other agents. A short analysis of the results is then presented.

2.1 Shadoc: A Conscious Co-ordination of Actions

Context. The first model, Shadoc, represents a community of farmers using an irrigated scheme [1]. It was built so that to investigate why most irrigated systems in the Senegal River middle valley are underused relative to their design capacities. An irrigated system, as it has been observed during the field study that led to building of the model, is a network of canals that bring water from a source to plots for agricultural production by farmers. These farmers are grouped into producer organizations, and trade between themselves and with other people out of the system; the whole may or may not be managed by a users' association.

There are different levels of complexity in the description of the system. First, the ecological system in itself is not out of one piece: the water that is pumped from the river has to be taken in charge by a lot of different actors along the time. Then the relationship among the different actors are not simple: each one is included into different groups, interconnected through membership and service relations and in addition to this complexity due to their nature, these systems are also the reflection of complex social relationships. Complexity also arises out of the need for some actors to wait until others have performed a particular task before they can take action. For example, a farmer must wait until the pumps are turned on before he can start work in this plot, but, in some systems, the pumps are not turned on until enough farmers are ready to begin.

Agents and Representations. This model is based upon certain assumptions. Basically we assume that irrigated systems may be represented as a place of acquisition and distribution of two resources: water and credit. The PumpStation allows the access to the shared good, water. The quantity of water depends on the functioning state of that PumpStation and on the choice of operating rule made by the agents. The agents in the model stand for different organisational levels: some are individual actors (farmers) and others are groups [17]. Simulated society is assumed to be structured with only three types of irrigated system related groups described below, one friendship network, the archetype of all existing networks, and a four-level social categorization.

Each farmer agent belongs to three different kinds of group agent:

- one whose grounds attribute is #waterAllocation among farmer agents with their plot attribute along the same Watercourse instance,
- one whose grounds attribute is #pumpStation management. This group agent includes all farmer agents of the scheme represented,
- one whose grounds attribute is #creditAccess for its members, which composition is *a priori* different to the composition of #waterAllocation group agents.

Each farmer agent acts also according to his own goal attribute from the set {#production, #food, #land, #waitingBetter}. Each agent has also its own point of view about the state of the system and especially its potential relations with other agents. This representation contains his knowledge of the water level on his plot, his friendship network constituted by other farmer agents with which it will be willing to cooperate through exchange of information or services and finally assumed rules enforced by agents with which it is in relation.

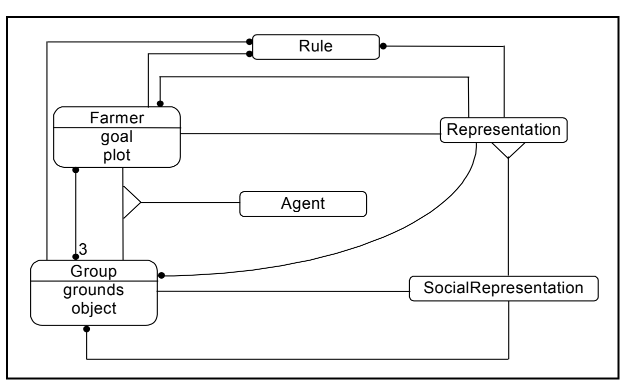


Fig. 1. Social structure of the Shadoc model

The Farmer and Group classes both inherit from the Agent class. Each Farmer class instance is in relation with exactly three instances of the Group class, one instance of the Representation class (which inherits from the SocialRepresentation class) and several Rule class instances. Each Group class instance is in relation with several Farmer class instances, several Rule class instances and exactly one SocialRepresentation class instance. Each Representation class instance is in relation with several Rule class instances, several Farmer class instances and several Group class instances.

Plot sowing divides the irrigation stage into two steps for each farmer. Initially, the purpose is to get information about the start-up of the pump station, either from a farmer in the friendship network or from the pumping grounds group. When it gets this information, it updates his representation of pump station functioning state and seeks for sowing his plot. Then, at each time step, the farmer-agent has to choose to go or not to go to the field, depending on its goal and its representation of the plot water level and pipe state, pump station functioning state and the possibility of getting water. At the same time, pump station management group updates at each time step its representation of irrigated scheme state and, according to the fulfilment of its rule of ending cropping season, it uses its representation to stop pump station. Then each agent switches to third stage: each one evaluates its satisfaction criterion. If satisfied, it changes nothing in its rules. If not satisfied, it may change its goal and its set of rules according to one of the following methods:

- changes nothing,
- imitates the agent of its "imitation group" with better score for its own criterion,
- selects agents of its "imitation group" who would have satisfied its criterion and imitates the one with the behaviour nearest to its own,
- imitates the agents who got the best result for a fixed time and then reverts to his former behaviour if not satisfied.

For each farmer, the "imitation group" is the group of farmers from which it may collect information (also called in this version of model its “friendship network”) and to which he thinks it is worthwhile to compare itself. This may take different forms for each farmer:

- The whole friendship network,
- the intersection of the friendship network with the water allocation group,
- the intersection of the friendship network with the credit access group,
- the intersection of friendship network and water allocation group and credit access group.

For each group, the "imitation group" is the set of all other groups with same grounds. In imitating another agent, any agent adopts the whole set of rules of this agent for itself.

Simulations: Influence of Social Networks on Irrigation Schemes. The influence of social networks on the viability of irrigated schemes has been tested. A simulation is considered to be viable if it lasts more than four cropping seasons with at least one cropping season after the fifth one practiced by more than 20 % of Farmer agents. There are several varying parameters:

- "homogeneity". If this parameter has a "true" value, #creditAccess group agents and friendship networks have the same composition as #waterAllocation group agents. Otherwise these group agents are *a priori* independent,
- number of friendship networks,
- vector with percentage of farmer agents for each mode of imitation set constitution.

What the results show is that even if social networks cannot be used to explain the viability of some scenarios, the number of friendship networks still has an effect: the less divided the population, the more viable the irrigated system. That means, unsurprisingly, that for one scenario, the reduction in the number of friendship networks raises the probability of viability of a simulation, in other words, the more people get along the higher the probability of viability of a simulation for a given scenario.

A population of farmers with imitation sets limited to farmers along the same watercourse seems here to be another factor raising the probability of viability for a simulation of a given scenario.

Finally, the homogeneity parameter seems to have no effect according to Table 1 which compares mean frequencies of viable simulations for each homogeneity value, the number of friendship groups being fixed at 5 and imitation sets varying.

Conclusion. This first experiment shows that viability of irrigation systems is enhanced when more people accept to share their knowledge about this system. Furthermore: the more these people are connected through the resource, the greater the effect.

2.2 JuMel: Emergence of Heterogeneity

Context. The model was built as part of a study on cattle transhumance in the extreme North of Cameroon. In the soudanian zone, the resource is highly irregular and herdsmen have to travel with their cows to feed them easily, sometimes for long time-periods. Thus, they leave their usual homes and perform movements that are called transhumance. Although the land seems to be open [13] there exist unwritten rules that define how these "commons" should be used [15]. During a field study, an economic analysis was aimed at finding these rules. It showed that, at one moment, the choices were based as much on the individual relationships with local people as on the grazing patterns. More specifically, it was shown that these relationships show great regularity [16]. To interpret this regularity, it is possible to refer to the different economic models: some consider that any action can be interpreted as a cost, and that regularity, creating familiarity, is a good way to reduce transaction costs [14], [21]. But it is also possible to regard the relation as an element that is important in itself, and that the agents want to preserve the history of their interactions thanks to that regularity [11]. The image that each one has of the others, the way it is built and how it influences the actions, seemed to be an interesting way to question the economic habits of the transhumant people [18], [19].

The aim of our research was to analyse the use of the resources and the patterns of relations that could be created simply through merchant exchanges, in an artificial world that captures the situation of herdsmen trying to have access to a resource over time. To build the agents' logic, the inspiration came from the economic points of view previously described, and two ways of reasoning were implemented for our agents: one inspired by the idea of cost, and the other by the idea of institutions.

Since we wanted to generate regularities in relationships, the choice was to decide that the farmer who expressed a refusal would be made somehow unattractive. This was translated, using no more than the minimum cognitive processes:

- To capture the idea of costs: the herdsman integrates any refusal as an extra-cost anticipated for the new transaction.
- To capture the idea of a value in the relation: the herdsman remembers the refusals and the agreements made by the farmers.

Agents and Representations. We put together three classes of agents (Figure 2). The first population represents herdsmen who need the grass and water and request access by making proposals. The second represents the village chiefs who accept the presence of the herdsmen by providing good or poor access to water depending on their order of arrival. The third is that of the farmers who are grouped together in villages under the responsibility of the village chiefs and who own land that they may allow herdsmen to graze.

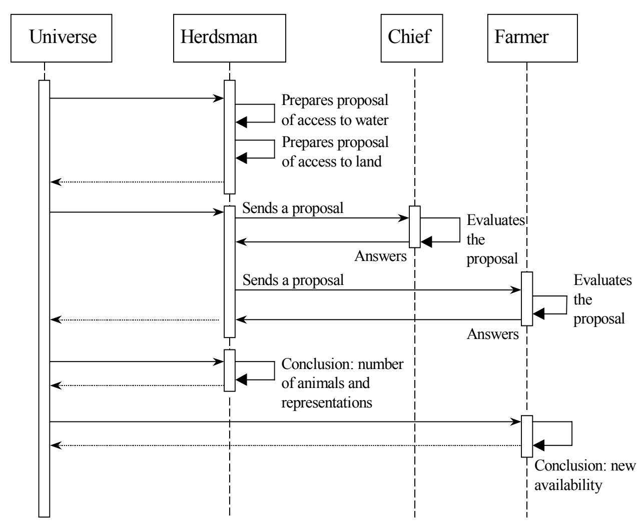


Fig. 2. A simulation step

The representation is based on the past exchanges between the herdsmen and the farmers. A farmer always charges the same fee when he allows a herdsman onto his land. A village always charges the same amount for a good site but accepts whatever the herdsman offers for a poor site. To start with, the herdsmen do not know the value of any of the access fees. When a farmer accepts an agreement or when a village provides access to a good site, the herdsman registers the cost of the transaction as a representation of the other agent. Any time the herder receives a refusal from a farmer or when a village provides access to a poor site, the representation (the access fee he is willing to pay) changes: it is increased by the « learning constant ». The herdsmen also remember the "quality of the relation" which is:

number of good accesses – number of bad accesses (for a chief)

number of proposal accepted – number of refusals (for a farmer)

In the experiments described here, there are two possibilities available to the herdsmen for choosing which village chiefs and farmers to ask for access:

- The first option is to choose the « cheapest » village, and address a demand to its chief. Since the agent will have to pay for the fee of the chief and for two other agents, the imagined cost for a village has a value of:

  Village price = the chief fee + 2 *(average of representations of known farmers in the village).

  Once the villages with the lowest fees are chosen, then the herdsmen choose the cheapest farmers in each village. As the herdsmen do not know the fees *a priori*, when they make this choice they can only ask the agents that they already know.

- The second option is to choose the village where he had the largest excess of compared over refusals. The global indicator for the village is:

  Village's indicator = ("quality of relation" with the chief) + (average of "quality of relation" with the farmers of the village)

Then he addresses the chief, and in the village he asks the farmers that have the highest quality of relation.

Conclusion. In the results (Figure 3) one notices that the resource is highly depleted, and then the herds are quite small when the agents choose with the "cost priority" choice, whereas it is almost at its best when then choose with "friend priority". Indeed, the "cost priority" representation creates a very important competition among the herdsmen.

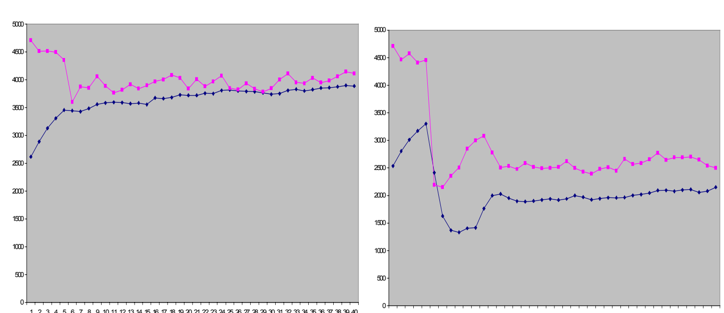


Fig. 3. Total number of animals in the environment and total number that could be accepted as a maximum by the resource along the time (400 time steps). On the left, the choice is made with "friend priority": the resource is not too depleted compared to the beginning, and is quite well used. On the right, the choice is made with "cost priority", which is bad for the resource, and where it is not so well used

One can see that usually in the "friend priority" choices, each transhumant has a different way of classifying the agents that are the most interesting to it. That difference of classification can be explained by the fact that the agents' representations are based on individual criteria (the quality of the relationship) in the "friend priority" cases whereas they are based on objective criteria (real prices) in the "cost priority" ones. Indeed, with the "friend priority" mechanism, the agents remember the history of their encounters, which unable them to develop representation that are all original; on the opposite, with "cost priority", they have more chance to share interest for the same farmers.

If one observes the error of the agents, this error being for one agent the sum of the differences between its representation of the price and the real price for all farmers he knows. In the cost priority simulations, this error is what creates the individual representation of the agents since it is the only difference of representation between all of them. One observes then in "cost priority" simulations, when one varies different initial parameters, that there is a high correlation between the mistakes that the agents make in the evaluation of the prices and the survival: the highest the mistake, the better the results.

That correlation of "having heterogeneous representation" and having a good use of the resource is just related to the definition of the access. In the model, a limited number of animals can be accepted on the fields. The results show that the resource is much less depleted in that case if not all the agents regard the same places as attractive, like in the "cost priority" simulations. If all transhumant are tempted to ask to the same farmers, only a few of them won't be refused and the resource will have a high chance to receive too many animals.

Fig. 4. Locations of the traps' paths on the spatial grid for experiments R1 and R2. The first layer, (slight grey), represents standard cells. The second layer is the presence of water (medium grey), the third one is the presence of road (in black), and the last one (dark grey) is the presence of traps

In our study, having heterogeneous individual representation to perform action is what allows the agents not to create competition that depletes the resource.

2.3 Spatial Representation and Groups Coordination: Djemiong Model

Context. This example is based on a study held in an eastern Cameroonian forest village named Djemiong. In that area where there exists no protected area, the challenge is to understand how the resource is managed locally. The aim of the research is to elaborate a model in order to study the viability of this management scheme. The major hunted specie is the blue duiker (*Cephalophus monticola*, Thunberg). Surveys have been conducted to understand the hunting behaviour of the inhabitants [20].

In the Djemiong village, the resource is hunted six months/year and there is a spatial shifting rule. Each year each hunter changes the location of his traps. The hunters present this behaviour as a management rule. Can the rules regulating the access to space at different moments of the year be considered as a management rule? This is the key-question of this study. Consequently, we need a spatial and dynamic simulation of the resource component, but we also need to take into account the behaviour of the hunters, the way they interact when they decide to locate their traps on the village's territory.

Data has been collected to simulate the life history of the blue duiker. A duiker agent has been created. Its attributes are the age, the sex, the gestation length and the partner. The behaviour of the blue duiker is implemented through one method that uses the above life-history parameters to define, with a weekly time-step, the growth, mortality, and reproduction functions, and also some rules for the movements [4].

Agents and Representations. 90 hunters agents have been created. By simulation we first studied the scale of the agents' spatial representation: their representation concerns the space, which is represented at various scales. The lower scale is the elementary cell of the resource model. The intermediate level is the path level: each agents knows the positions of trap paths which is composed of eight elementary cells. The upper level is an area limited by rivers and roads. In a first set of simulation it has been demonstrated that the relevant representation for the agents is the intermediate level, the traps path.

We also simulated the co-ordination of the agents: each agent has a collection of 4 precise traps' paths, which is randomly initialised from a set of 4 hunting localities. A first scenario, denoted H1, states that a hunter agent chooses its current traps' path randomly among the 3 one he was not using during the previous season. This scenario stands for individual turnovers without any coordination between the hunters. To account for such a process, a second experiment (H2) defines 30 groups of 3 hunter agents. The 4 precise traps' paths are assigned to these collective entities that may represent kinship networks of hunters from the same family. The individual turnovers' rule also applies here, but with an additional constraint: a traps' path remains not hunted by any of the 3 group's members during 3 successive seasons.

Table 1. Global results for the second set of experiments (with hunter-agents)

Experiments	Hunting coverage (cells' number)	Population density after 25 years (number of animals/km^2)	Total catches during 25 years (number of animals)
H1 – Hunters' periodic individual turnovers	640.8 (28.4)*	15.77 (3.4)	12073.7 (766.1)
H2 – Hunters' periodic collective turnovers	747.2 (26.5)*	24.21 (3.4)	13970.3 (824.3)

*Calculations based on 10 simulations. Absolute frequencies and standard deviation is given in brackets.

For the second set of experiment, we can see that the H1 scenario is the worse: after 25 years, the population and the captures are still decreasing. The low value for the hunting coverage (table 4) suggests that there is probably a problem of spatial congestion (90 * 4 = 360 individual traps' paths have to be defined, some of them should overlap, and then all the hunter agents could not access to their path if it is already occupied by others). There is probably another reason to explain the difference of about 9 animals per Km^2 with the population density of H2 (table 4). The lack of co-ordination between the hunter agents for scenario H1 should erase the effect of the individual turnovers.

Conclusion. In Djemiong village it has been observed that places are inherited in the families. By inheriting parts of the village area from their fathers, hunters take their decisions of traps locations on a partial representation. Thus, the management of the resource is more complex than imagined: it has to be understood through the links between social organisation and spatial structure. Dealing with that complexity, multi-agent simulation is used to identify the relevant organisation level (groups rather than individuals), the relevant spatial scale (path rather than hunting localities) and how the agents co-ordinate to avoid competition.

2.4 Spatial Representation and Collective Decision: CommonForest

Context. This model is a theoretical one that tends to question issues from the field of geography [2]. In this model, agents use a resource, and for that can take collective decisions so that not to deplete it. Herder agents will forage in a fragmented landscape composed of several forests. A model of forest diffusion provokes an increase of the forest cover. The herds consume the forested cells of the spatial grid. Agents have individual representations of the space they share and act according to that representation.

Agents and Representations. The herder agent memorises each cell it has foraged and the state of the cell at that moment. At any time step the agent uses a specific threshold to decide if it grazes or not: if the group of forest cell is bigger than the threshold it uses, it is free to use the forest. The individual representation of space is simple: rather than a mental map composed of a visited cells array, the representation is the sum of states of forest cells they have encountered. It helps to decide of the individual threshold: the more forest cells the agent meets, the more abundant the forest seems to be and the lower the threshold. That representation is periodically compared with others' perceptions to build a collective threshold that is computed at the village level. The village entity is the group of agents that defines a collective threshold to constrain the consumption. For that, the forests which size is under the collective threshold will be protected against consumption.

Three strategies are defined.

- The first one is called individual strategy. The agent does not take into account the collective constraint and acts according to its individual representation.
- The second one is called collective strategy. The agent will respect the rules imposed by the group and avoid the protected forests.
- The third strategy is a mixed strategy. In case of conflict between the individual and the collective representation, the agent will compute a mean threshold.

The goal of the simulations is to compare the three strategies. The initial state of the landscape is presented in figure 5. Periodically the contiguous cells are aggregated in a forest entity. Depending on the collective threshold this forest is marked as protected or not. In that case the cells will be protected against the strategy 2 agents. With the strategy 1 and strategy 3, agents decide by themselves if they consume that resource.

Before the comparison of the three strategies we have tested a scenario without any representation. As a result the forest always disappears between time step 150 and time step 250.

The first scenario tests the individual strategy. Each herder agent will decide which forest it will consume depending on its past history. The maximum value of the threshold is 10. For a period of ten time steps the agent will move and decrease its threshold each time it is placed on a forested cell. If the agent has encountered few forested cells in the past ten time-steps, its individual threshold will be high and thus larger forests will be protected.

The second scenario tests the collective strategy and the third scenario tests the mixed strategy.

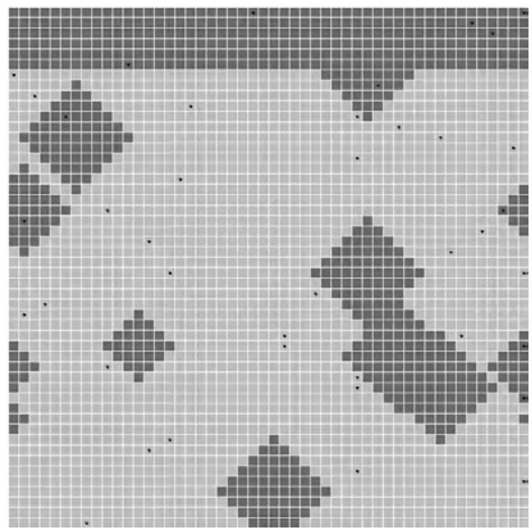


Fig. 5: initial state. The forest is represented in dark grey (686 cells grouped into 11 forests). Herder- agents are represented by a dot. The whole landscape is made of 50*50 = 2500 cells with closed boundaries and Moore neighbourhood

Table 2. Simulation results. These results are given for 300 time steps simulations

Strategy	*Forested cells*	*Number of forests*	*Mean threshold*
Individual	*150*	*20*	*8*
Collective	*230*	*61*	*5*
Mixed	*233*	*63*	*6*

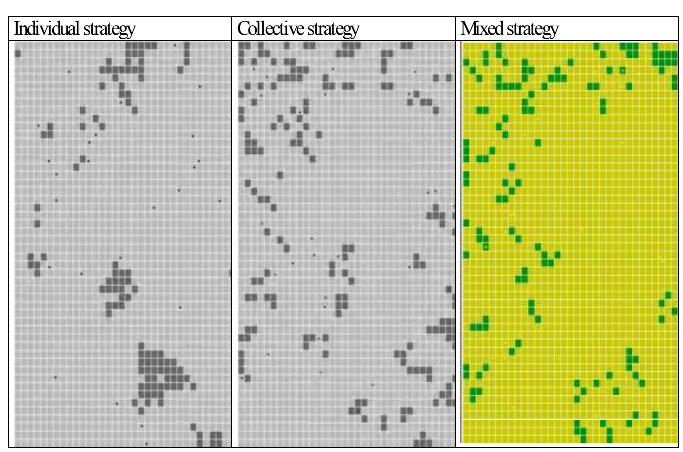


Fig. 6: Landscape state after 300 time steps

One can observe the individual strategy gives the worse results both in terms of resource preservation (the forested cells indicator is the worse) and in terms of agent satisfaction. The threshold gives a good measure of the agents' satisfaction. The less the agents finds resource the highest its threshold is. The landscape state (figure 6) is also very different. The individual strategy leads to larger forests and the collective norm leads to more fragmented landscapes with more resources. Figure 7 shows that the scenario with collective decisions shows the greatest heterogeneity of the representations. The interpretation is that the distribution of the resources leads to diverse local histories of the agents. While foraging some succeed and some fail. The individual scenario is more homogenous, probably because of the lack of resource. There are minor differences between the collective strategy and the mixed strategy. By repeating the simulations, we observe that the mixed strategy gives more variability in the results.

Conclusion. The purpose was here a theoretical one: we wanted to study the links between a dynamical landscape and the actions of the agents that depend on their representations. In that exercise, they can act accordingly to a common norm, which is called a common threshold or according to their individual threshold. In any case they keep their own individual point of view on the resource. What is shown is that as long as it comes to protection of the resource, the result is much better when the agents share their criterion of the choice of protecting or using the resource. What appears however is that the heterogeneity of individual representations for the state of the forest is all the more important than the choice is common and the resource protected.

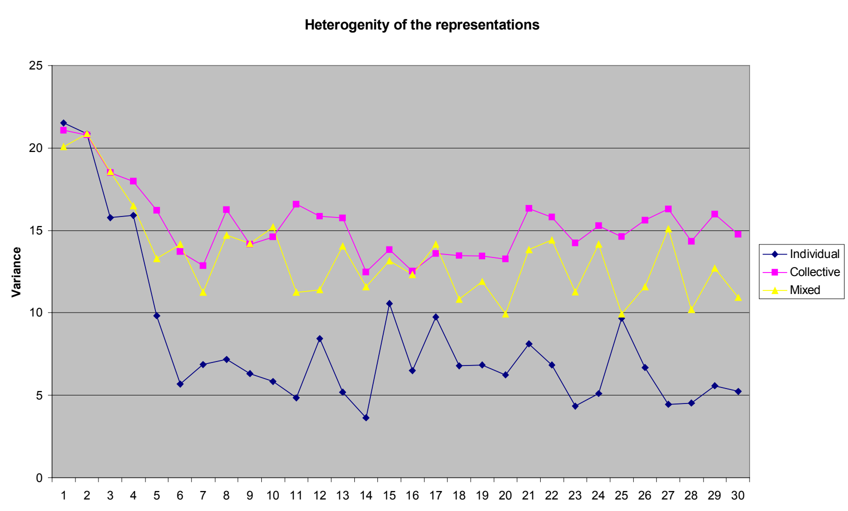


Figure 7: Heterogeneity of the representations. The values represent the inter-individual variability of threshold value. The difference is the most important when the individual all agree on the threshold for action and is equivalent when they have a mixed strategy. The representation is however much more homogenous when they all act in an individual way

3 Discussion

The four models presented here were created to tackle different case studies that deal with the use of renewable resources among agents. The representation formed by the agents about the resource reveals itself as important as we had imagined when deciding to focus on that aspect. What is pretty clear in our results is that competition has a bad impact on the production. This is an assumption that is often put forward by the people who try to create artificial societies [6]. But it is usually accepted that the agents have to agree on their representations so that one can observe a non-competitive behaviour [5], [12]. This is not necessarily what our results show.

In each case described the individual representation that exists for each agent is more or less related to the representation of the others. Homogeneity of representation can be good for the resource and the society of agents in some cases. This was very clear for one of the models, Shadoc, where the system would work better when the agents have homogenous representations of the actions to undertake. In another, CommonForest, if the agents agree on the representation they must choose for action, the resource is less depleted, but on the other hand the individual representation of the resource gets much more heterogeneous. For the other models, it is clearer: only a high difference of perception among the agents can secure the sustainability of the system. This appearance of contradiction is interesting since it questions the a priori put in the description of the systems.

In the Shadoc model all the agents should share a common knowledge and common rules so that the system can work better. Actually the resource is here described right away as a production tool where a very high degree of co-operation is needed. That co-operation is very simple, but it is anyway put into the system at first. As a direct consequence, the agents cannot all have the same activity and this forces them to agree at least partially on their representations. If there is not that minimum degree of agreement, the agents are not even able to use that resource.

In CommonForest, one sees that the homogeneity of actions is important. But actually, it is not a homogeneity of actions to perform that is fundamental, but the agreement on the actions that must not be performed. The common protection of certain places leaves the agents free to act in an independent way and they acquire, as a result, very different points of view on the resource. The more they agree on action, the more they disagree on their perception.

In the two other models, the homogeneity of representation is damageable for the resource. It is true when the representation depicted is fixed (Djemiong) and embedded in the social structure as well as when it emerges from the interactions (Jumel). In these cases, the fact that agents disagree, or have a false representation of the environment is the very element that makes the system sustainable, and this is for cases that seem to be quite usual in the world: when resources can be obtained by any individual agent that arrives first. This can be somehow related to the idea that some misunderstanding can be a key element of the building of a social group [9].

To be able to synthesize these results, it seemed interesting to draw two categories for the resource. In some cases, it is interesting to point out that the agents need to co-ordinate to get goods and that the resource can be considered as produced by them in an active way: for that then they must gather at some occasion and it is better if they agree on some common representation of the world. In other cases, one can consider that the allocation depends on the repartition of the agents in space and requires an even distribution, and it is then better for them not to be interested by the same goods. In that case, we saw two good ways to picture it: either to give to the agents a representation that make them consider different spaces as good for them, or to make them agree on the places they should avoid.

What we conclude is that there is not necessarily one good solution to co-ordinate agents using a common good. The agents do not necessarily have to co-ordinate when they share the same place and/or the same resource: the need of coordination depends directly on the way the system is built and on the type of resource that one wants to describe. Before a choice of representation has to be made the very dynamics of action and resources are to be considered. Some misunderstanding among the agents can reveal itself very useful for the system.

This can be applied to the analysis of real systems (like in our simulation studies), but can reveal itself interesting for all people who have to create a system and organise the agents. If some elements that have to be used will be limited in the system, the way that resource will have to be implemented can be thought as competitive (or not), in which case the system would be optimal with agents that are heterogeneous in their representations.

Acknowledgements

Cirad supported the building of these models and the simulations that were led.

References

1. Barreteau, O., Bousquet, F.: Shadoc: a multi-agent system to tackle viability of irrigated system. In: Annals of Operation research, Vol. 94 (2000) 139-162
2. Bonnefoy, J.L., Le Page, C., Rouchier, J., Bousquet, F.: Modelling spatial practices and social representations of space using multi-agent systems. In: Ballot, G., Weisbuch, G (eds) Applications of simulation to social sciences. Hermes (2000)
3. Bousquet, F.: Distributed artificial intelligence and object-oriented modelling of a fishery. In: Mathematical Computer Modelling, Vol. 2018 (1994) 97-107
4. Bousquet F., Le Page C., Bakam I., Takforyan A: A spatially explicit individual-based model of blue duikers population dynamics: multi-agent simulation of bushmeat hunting in an eastern cameroonian village. In: Ecological Modelling, in press (2000)
5. Conte, R., Castelfranchi, C.: Mind is not enough : the precognitive bases of social interaction. In: Gilbert, N., DOran, J.(eds), Simulating societies, UCL Press (1994) 267-286
6. Conte, R., Gilbert, N.: Computer Simulation for Social Theory. In: Conte, R., Gilbert, N. (eds) Artificial Societies. The Computer Simulation of Social Life, UCL Press, London (1995) 1-15

7. Conte, R., Castelfranchi C.: Distributed artificial intelligence and social science : critical issues. In: O'Hare, G.M.P., Jennings, N.R. (eds.), Foundations in Distributed Artificial Intelligence, Wiley Intersciences Publications (1996) 527-542
8. Costanza R.: What is ecological economics ?. In: Hall, C.A.S. (ed) Maximum power, the ideas and applications of H.T. Odum, University Press of Colorado (1995) 161-163
9. Doran, J.: Simulating Collective Misbelief, In: Journal of Artificial Societies and Social Simulation, Vol. 1, no. 1 (1998)
10. Folse, L., Packard J., Grant W.: AI modelling of animal movements in a heterogenous habitat. Ecological modelling. Vol. 46 (1989) 57-72
11. Greif, A.: Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders. In: The journal of economic history, Vol XLIX, no. 4 (1989) 857-882
12. Haddadi A., Sundermeyer K.: Belief-Desire-Intention Agent architectures. In: O'Hare and Jennings (eds), Foundations of Distributed Artificial Intelligence, Wiley InterSciences (1996)
13. Hardin, G: The tragedy of the commons. In: Science, Vol. 162 (1968) 1243-1248
14. North, D.C.: Institutions, institutional change and economic performance, Cambridge University Press, Cambridge (1990)
15. Ostrom, E.: Governing the commons. The evolution of Institutions for collective action. Cambridge University Press, Cambridge (1990)
16. Requier-Desjardins, M.: L'accès aux pâturages, une approche économique de la mobilité, In: Actes du colloque Méga-Tchad, L'homme et l'animal dans le bassin du lac Tchad. Orstom, Paris (1997)
17. Rouchier, J., Barreteau, O., Bousquet, F.: Evolution and Coevolution of Individuals and Groups. In: Yves Demazeau (ed), Proceedings of the Third International Conference on Multi-Agent Systems, IEEE Los Alamitos, USA (1998) 254-260
18. Rouchier, J., Requier-Desjardins, M.: L'interdisciplinarité pour la modélisation dans la recherche-développement. Compte-rendu d'une experience en cours, une application au pastoralisme soudano-sahélien. Proceedings of SMAGET, Cemagref, Clermont-Ferrand (1998) 193-204
19. Rouchier, J., Bousquet, F., Requier-Desjardins, M., Antona, M.: A Multi-Agent model for describing transhumance in North Cameroon: comparison of different rationality to develop a routine. In: Journal for Economic Dynamic and Control, Vol. 25 (2001) 527-559
20. Takforyan, A.: La chasse : gestion communautaire et logique économique (Cameroun). In: Compagnon, D., Constantin, F. (eds), Karthala, Paris, Ifra, Nairobi (2000) 155-177
21. Williamson, O.E.: Les institutions de l'économie. Inter Editions, Paris (1994)

Understanding Climate Policy Using Participatory Agent-Based Social Simulation

Thomas E. Downing[1], Scott Moss[2] and Claudia Pahl-Wostl[3]

[1]Environmental Change Institute, University of Oxford,
1a Mansfield Road, Oxford OX1 3SZ, UK
tom.downing@eci.ox.ac.uk
Centre for Policy Modelling, Manchester Metropolitan University,
Manchester M1 3GH, UK
s.moss@mmu.ac.uk
Department of Systems Analysis, Integrated Assessment, and Modelling, EAWAG,
Überlandstrasse 133, Dübendorf CH-8600, Switzerland
Claudia.Pahl@eawag.ch

Abstract. Integrated assessment models (IAMs) have been widely applied to questions of climate change policy—such as the effects of abating greenhouse gas emissions, balancing impacts, adaptation and mitigation costs, understanding processes of adaptation, and evaluating the potential for technological solutions. In almost all cases, the social dimensions of climate policy are poorly represented. Econometric models look for efficient optimal solutions. Decision making perspectives might reflect broadscale cultural theory, but not the diversity of cognitive models in practice. Technological change is often ignored or exogenous, and without understanding of stakeholder strategies for innovation and diffusion. Policy measures are proposed from idealised perspectives, with little understanding of the constraints of individual decision makers. We suggest a set of criteria for IAMs that can be used to evaluate the choice and structure of models with respect to their suitability for understanding key climate change debates. The criteria are discussed for three classes of models—optimising econometric models, dynamic simulation models and a proposed agent-based strategy. A prototype agent-based IAM is reported to demonstrate the usefulness and power of the agent based approach and to indicate concretely how that approach meets the criteria for good IAMs and to complex social issues more generally.

1 Introduction

Climate change brings into sharp relief the most difficult problems in understanding the interactions between the natural environment and socio-economic conditions, and in addressing the integration of science and policy.

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 198-213, 2000.

Climate change is unusual, if not unique. Instead of trying to solve an existing environmental crisis, policies must prevent a problem that is only now emerging, at least according to some scientists [7]. The processes leading to climate change are not well understood. A validated theory of the whole of the relevant physical system is lacking [18]. Many social theories are relevant and none are entirely adequate; yet some respected scientists argue strongly that climate change is caused by the behaviour of social systems [14]. Both natural and social systems are characterised by enormous complexity and the interaction between those systems adds further complexity. Moreover, the information available to policy makers is plagued by large inherent uncertainties due both to measurement errors and conceptual ambiguities.

Understanding and predicting climate change—including the evolution of the climate system, impacts and human responses—is beyond present and any foreseeable scientific capacity. The normal approaches of the natural sciences, direct experimentation and building predictive models, are not applicable to environmental threats on the scale of global climate change.

A new basis for devising public policy is required.

In this paper, we

- Review the characteristics of climate change and the implications for policy analysis in integrated assessments.
- Evaluate the policy requirements and the merits of alternative methodological approaches for understanding climate policy. We focus in particular on two genres of IA and compare these with agent-based social simulation.
- Illustrate the design of an agent-based approach with a pilot model, including specific policy questions and the relevance of the proposed methodology.
- Discuss methodological issues related to agent specification and interactions for models of such complex environments.
- Conclude with general observations on the role of modelling in integrated assessment, and climate policy more generally.

The pathway we propose begins with software representations of real actors. Agent based social simulation models are concerned with the ways in which social structures emerge from interactions among individuals and how those structures influence and constrain individual behaviour, thereby altering or reinforcing social structures.

The advantage of agent based social simulation is that it combines the problem orientation and commitment to observation of the sociologist and anthropologist with the more formal approaches of the natural scientist. We argue that agent based social simulation supports a new methodology that itself provides a suitable framework within which to collect observations of social and physical systems, to generalise from those observations and to identify relationships and processes that must be understood to shed insight into policies to deal with climate change.

2 Integrated Assessment of Policies for Climate Change

Integrated assessment (IA) is “a structured process of dealing with complex issues, using knowledge from various scientific disciplines and/or stakeholders, such that integrated insights are made available to decision-makers" [15]. In the lexicon of integrated assessment [12, 17], the following features are highlighted:

- Use of models to structure a complex environmental issue.
- Participatory methods to define the problem, choose methods and interpret results.
- Interactive exercises with stakeholders, experts and models to shed insight into the nature of the problem and policy options.
- Methodological concerns with uncertainty and robust results, including sensitivity to temporal and spatial scales

In many respects IA is not new, but it is an effort to develop robust approaches to complex environmental problems. We differentiate between IA as a process of analysis and stakeholder engagement and the Integrated Assessment Models (IAMs) that are often more visible than the process.

Regarding climate policy, integrated assessment must span greenhouse gas emissions, effects on climate and sea level, impacts on managed resources and unmanaged ecosystems, and interactions with human social and economic activities. Each of these is a major undertaking.

Two broad types of IAMs have dominated the field:

1. Econometric models based on economic theory and calibrated to recent experience address near-term mitigation policies. Policy models in this arena began with computable general equilibrium models (e.g., DICE [11]), and subsequently sectoral, bottom-up simulation models (such as the energy-environment-economy models from Cambridge Econometrics).
2. Dynamic simulation using more or less linear couplings of representations of each sub-system focus on sustainability, the carbon cycles and long-term environmental futures. Two prominent examples are Targets [16] and Image [1]

Neither of these approaches achieves the level of integration and policy analysis that is required.

The challenges of climate change policy analysis for IA include the following (see Fig. 1):

- Climate change affects almost all natural systems and human socio-economic activities:
- – Climate change policy needs to draw on many disciplines across the natural and social sciences.
- – Key sectors must be modelled at acceptable resolutions, representing the critical processes that influence the rest of the system. A common critique, for example, is that most IAMs use very simplistic climate models.
- Climate change is global:
- – Spatial resolution must be realistic. While a full understanding of every patch on the Earth's surface (not to mention the oceans) may not be required, a world model based on a dozen regions is certain to ignore important issues.
- – The global and local must be joined, in terms of policy and response, impacts and consequences, and globalisation of societies and economies.
- The time scales of climate change range from the immediate to the centennial or longer:

– Fundamental questions about intergenerational responsibilities must be included.
– The future may be best represented as a succession of decision nodes, comprising different pathways.
– Technological and social change will be substantial on the time scales of climate change. Understanding how and under what conditions long-term change occurs is critical.
• Climate change poses a threat to Earth's sustainability, as well as opportunities for new resources:
– Visions for the future must be plausible, desirable and realistic. The target for policy should not be constrained to simple balances, such as cost-benefit, economic welfare, or CO_2 stabilisation. And, undesirable futures must be realistically portrayed and assessed. Preventing an undesirable future may be easier than reaching consensus on how the world-system ought to evolve.
– Climate change issues must be embedded within broader social goals of development, sustainability and equity (*e.g.*, [8]).
• Climate change is uncertain in almost all of its manifestations:
– Large, apparent impacts will spur responses, but detection and attribution of underlying changes, particularly at the local level is difficult. The local signal of climate change is weak compared to the noise.
– Optimality may be impossible; actions must be portrayed as an evaluation of risks while learning more about the threats.
• Climate change is global in the sense that it is the manifestation of diverse human activities and the actions of multiple stakeholders:
– Conflict between the aims of different stakeholder should be explicit. The domain of climate policy is inherently one of conflict—we should not presume consensus and voluntary adoption of recommended policies.
– Assessments should not portray the world as subject to a single decision maker or decision making paradigm, whether optimising or not.
– To achieve policy insight, modelling must engage stakeholders.

Clearly, global change issues require a new science that addresses these fundamental requirements [14].

This rather long list of 'desirables' underpins our analysis in the next section of two prevailing genres of IAMs and the promise of participatory agent based social simulation in climate change and policy analysis.

3 Alternative Approaches to Integrated Assessment Modelling

The major strategic challenge in integrated assessment is ensuring that policy insights from modelling are robust. This requires that the modelling is appropriate to the policy issue. It is almost never the case that a single model will be adequate for all issues. Conversely, it is often the case that models are inappropriately applied to policy issues for which the design and structure are not adequate. And all too often, the structural suitability of different model types is hidden from the policy maker.

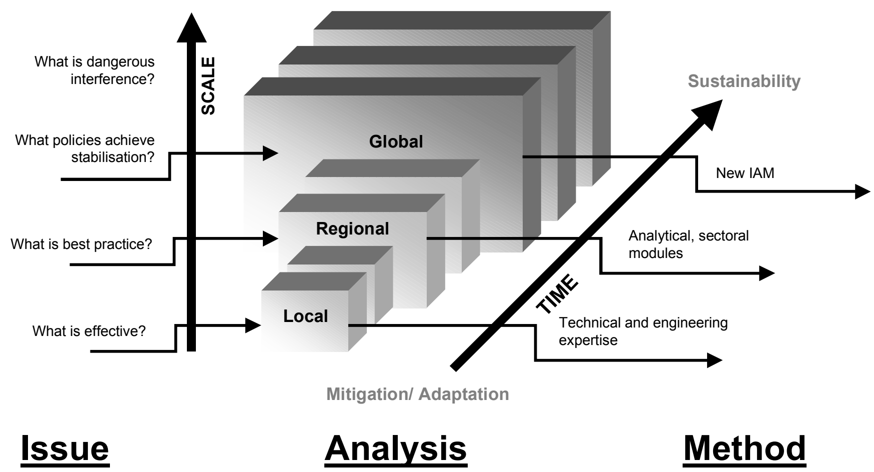


Fig. 1. The issues that have to be addressed in the integrated assessment of climate change pose formidable challenges across both spatial and temporal scales. These issues have been considered in, for example, [6, 7, 13, 14 ,19]

We identify 11 aspects of integrated assessment modelling that are often cited as essential. We compare these criteria for the two dominant modes of IAMs: the equilibrium, optimising policy models and the dynamic simulation models of Earth's sustainability. The classic example of the first, which tends to be relatively simple in terms of structure and resolution, is the DICE model [11]. The dynamic simulation models have a longer-term perspective, suitable for addressing the impacts of climate change and carbon cycle issues, among others. Classic examples are the IMAGE model [1] for a spatially explicit approach and Targets [16] for a more highly integrated model but with less spatial realism.

While there are many variants of IAMs. representing the field with two classic examples highlights the conditions in which such models might be applicable. Meeting these conditions of application are fundamental challenges to all policy modelling. We simply propose, and offer an early demonstration, that an agent-based approach is a new paradigm for the integrated assessment community that offers a significant source of insight into climate policy.

3.1 Model Specification Issues

Natural Systems. An effective IAM will represent processes in environmental systems that cause climate change or are affected by its consequences. The DICE model and others of its type are generally limited in this regard with restricted interactions between agents and the representation of the natural system. Dynamic simulation models commonly integrate representations of the natural systems at varying resolutions. The representations are always more highly articulated than in the DICE-type models and entail some degree of spatial realism. Whereas the DICE-type models presume that processes take place within each time step so that the observed properties of the model are equilibrium properties, the dynamic simulation models are process oriented and, so, do not require the assumption that equilibrium prevails.

Prototype ABSS models as described below implement environmental models of varying degrees of complexity and agents that seek to manage the environment, to mitigate the effects of environmental events and that react to environmental changes. The construction of these models allows for links to large-scale models of the environment.

Interaction among Stakeholders. The DICE-type models do not represent stakeholders explicitly. There is usually a single global policy maker, there may be a representative household representing all populations by means of a single utility function and sometimes a discredited kind of global production function.[1] The dynamic simulation models do not represent individuals explicitly but rely on sociological theories and taxonomies to identify patterns of response and favoured types of policies in the face of climate change. ABSS models are distinct from both of these types of existing IAMs in that investigating the consequences of social interaction among heterogeneous agents is the point of social simulation.

Agent Motivation. The DICE-type models all assume that social processes result from utility maximisation, constrained by incomes and prices, with no effects on consumption patterns or magnitudes as a result of social interaction. The dynamic simulation models are not usually explicit because they do not represent individuals directly. In ABSS models, the agents represent individuals or collections of individuals using a wide range of representations of cognition. All of these representations have in common an acceptance of bounded rationality and the behaviour of the agents can be compared with the observed behaviour of the social entities they represent.

Societal Structure. This is a key issue for the analysis of issues since it concerns the ability of fundamental socio-economic structures to change over time as well as the differentiation of social structures in different cultures. These issues cannot be incorporated into the econometric-optimisation models of the DICE type which are static, calibrated to past performance, ignore informal economic behaviour and cannot represent the sorts of surprises that lead to changing institutional structures. The dynamic simulation models can accommodate structural change which is usually related to external driving forces or presumed macro-level relationships (e.g., between GDP and governance). Such changes do not emerge from stakeholder behaviour and social interaction as they do in ABSS models.

Technology. The importance of technology in integrated assessment is due to the importance of innovation and the diffusion of innovations for the mitigation of climate change (*e.g.* by reducing emissions of greenhouse gases) and any direct impacts of technology on the environment. In the DICE-type models, technological changes are represented by functions which are known empirically to rely on time patterns of income distribution [*sic*] or they are represented as changing values of coefficients that are difficult to estimate. In dynamic simulation models,

[1] For references and a detailed justification of this claim, see Moss (1999).

technological change is represented by, for example, learning curves that have some assumed or estimated parametric values. The problem here is the importance of modelling technological change as a response to environmental change. Any such modelling will require the process of innovation to be captured. There have been some limited ABSS models of such processes (e.g. [9]), but this is an area requiring substantial development.

Resolution and Scaling. The grain of representation of social entities as well as the geographic and temporal resolution are issues on which the IAM community is currently focusing. [2] The optimisation/econometric models are always specified at a single scale, often representing the world in fixed regions (sometimes a single region, but up to a dozen is common). The dynamic simulation models are also single scale, often with a geographic layer (*e.g.*, 0.5 ° resolution). For both types of model, the time step is generally annual and there are problems embedding the analysis in multi-scale assessments. Drawing on agent oriented software engineering and validation issues related to agent based social simulation modelling, the ability to demonstrate the consistency of models at different grain is an important research issue to the ABSS community.

Validation. Of the three types of models considered here, only agent based social simulation models offer any meaningful prospect of model validation. Both the optimisation/econometric models and the dynamic simulation models are parameterised against short runs of data without the degrees of freedom for validation against data held back from the data used to estimate the model parameters. In ABSS modelling, validation can include the assessment of the accuracy of agent specifications as representations of social entities – a means of validation that is not available to the other modelling approaches.

3.2 Model Use Issues

Stakeholder Participation. The involvement of stakeholders is an increasingly important element of integrated assessment that has led the IAM community to look towards the ABSS community for support. A natural point of entry for the stakeholders it the assessment by them of the accuracy with which they are represented by agents in terms of the agents' simulated behaviour as well as in the assessment of the plausibility of the results. It is essential for model-based stakeholder participation that the stakeholders be able to interrogate the models or model operators to ascertain why social outcomes have emerged and the conditions leading to observed agent behaviours.

Hypothesis Testing. The particular and ineluctable concern in testing hypotheses about the social effects and sources of climate change is the central importance of uncertainty. Hypotheses about specific outcomes in such complex and complexly interacting systems are as likely as hypotheses about the timing and magnitudes of earthquakes. Where hypotheses can be useful is in relation to processes – the

---

[2] For example, the European Forum on Integrated Environmental Assessment reviewed and debated approaches in a workshop on scaling in integrated assessment in July 2000.

equivalent of the theory of tectonic plates. The two prevailing approaches to IAM generate hypotheses by means of comparative static analyses of some policy measure *vis á vis* some baseline. Neither addresses uncertainty issues directly. ABSS models identify processes generating structural and social change and can be used to consider possible strategic effects of policies intended to mitigate or respond to the impacts of climate change. Uncertainty issues are naturally captured by ABSS models insofar as it arises from limited cognitive capacities of individuals (bounded rationality) in the context of social interaction.

Verification. Verification of the formal properties of integrated assessment models is not an issue for the IAM community. There is, however, some concern that representations of behaviour should be consistent with (or drawn from) other scientific disciplines such as psychology, economics or sociology and that representations of the natural system should be consistent with the relevant physics, biology and related sciences. The optimisation/econometric models have never been verified as being consistent with any independent theory and are not consistent with observation of behaviour. Representations of anthropogenic effects on the natural system are extremely brittle.[3] Dynamic simulation models frequently adopt representations of the natural systems that are directly verifiable in relation to quite fine grained models of those systems. However, the representation of stakeholders as "cultural types" rather than by agents with identifiable individual characteristics makes verification of social behaviour and interaction problematic.

Transparency. It is essential that stakeholders recognise the underlying assumptions that influence model results. In the integrated assessment field, the implausible assumptions of micro-economic theory constrain model insight into the likely behaviour of complex systems, which is perhaps why such assumptions are rarely portrayed to policy audiences. The dynamic simulation approach is based on plausible assumptions and it is both possible and common to signal the most important influences on outcomes through sensitivity testing. There is some concern, however, that large numbers of assumptions might mask critical structural issues. In agent based modelling, it is possible to separate out the behaviour of individual agents and develop more fine grained models of those agents in order to tease out whether the simplifying assumptions made at the coarser grain can be validated at finer grain – a process that can be continued until a grain of model is reached where the assumptions are either plausible to domain experts or verified with respect to independently validated principles of (for example) social psychology or cognitive science.

3.3 Conclusions for Model Specification and Use

We have presented a detailed view, drawn from the experience of social simulation and integrated assessment modellers, of three genres of models. We note five salient conclusions.

---

[3] See Moss, Pahl Wostl and Downing (2000) for references.

1. To seek to formulate optimal policy over a long time period and many geo-political regions is likely to be dangerous. The role of modelling should be to test plausible outcomes from different strategies. Such strategies should include learning and monitoring. Evaluation of simple policies (such as a carbon tax now) is only a partial analysis, and may be dangerous if it ignores major structural changes.
2. Models based on presumed theories of behaviour, whether economic or cultural [20] are problematic. No existing or foreseeable social theory will enable us to predict behaviour over the course of the next decade let alone the century time scale of climate change. Economic approaches begin with techniques that are remote from any reality that we observe. Following Edmonds [5], the models constructed by economists are not constrained by observation or by the results of any well-established science. This characteristic of economics-based integrated assessment models is in stark contrast with the approach of the natural sciences where biological and chemical models are constrained by well validated physical theories.
3. The representation of social agents is critical to understanding diverse pathways. Models, like decision makers, should focus on social processes, informed by independently validated principles of social and cognitive science to enable us to incorporate (conflict and negotiation as major elements in the determination of global change policy.
4. Policy strategies should be set in the context of diverse scenarios of socio-economic change, environmental conditions, societal values and governance. The goal of robust insight requires full evaluation of the fundamental structures of society. Technology will certainly change the future, but only in combination with social norms, economic values and political power.
5. Realism—encompassing validation and verification—is essential to engaging stakeholders in meaningful analysis of IAMs. Models should provide features that are recognisable in reality. They should reflect different mental models of stakeholders. And, they should be validated against observed data.

4 A Prototype Agent-Based IAM

Climate change gives rise to a wide range of issues resulting from its myriad consequences. The European project on Freshwater Integrated Resource Management with Agents (FIRMA) is concerned with the consequences of climate change for such water issues as drought, flood, changing use patterns due to irrigation needs and the like. The particular policy issue addressed by the model reported here is the effectiveness of exhortation in managing domestic water use during periods of drought. Though developed as a prototype for integrated assessment modelling with respect to the Thames region of England, the model is sufficiently coarse grained and abstract as to be applicable to other regions and issues.

This early application illustrates our focus on social processes and interactions, links to the episodic fluctuations in the environment, and validation by stakeholders and domain experts.

In the UK, exhortation was apparently effective during drought periods until the water providers were privatised, the water company managers were then awarded large remuneration packages and, in the midst of a drought and restrictions on water use with campaign exhorting households to conserve water, it became public knowledge that nearly half of mains water was lost through leakage *and* the water companies claimed it was uneconomic to repair the leaks. In effect, the public view of the water companies was based on a perception of greed: the companies required households to conserve water so that the companies did not have to repair leaky mains so that, in turn, the companies' managers could receive large increases in their remuneration. So it is clear that one issue is the reputation and public confidence in the water suppliers, environmental managers and government.. In addition, there are likely to be cultural issues. Whereas, at least formerly, exhortation was a useful and successful element in water demand management strategy in the United Kingdom, Dutch colleagues assure us that, in similar conditions, exhortation has been ineffective in the Netherlands.

Our prototype water demand model describes social relations that support the effectiveness of exhortation. These relations are neither necessary (there are doubtless other relationships that could be described and have the same effect) nor sufficient (since we know that cultural and political factors can and do predominate). The point is to describe a plausible mechanism that has been validated independently by another research community.

The key concept on which the model is built derives from the social psychologists' *consistency principle*. This principle, for which there is substantial descriptive evidence, has it that individuals tend to agree the most with those whom they like the best and tend to like best those with whom they agree the most. There is a strong correlation between shared attitudes and attractiveness [3, 4].

This principle was given expression in the Thames prototype model by specifying household-agents who could communicate directly with, and observe some behaviour of, a limited set of "neighbours". The household-agents were located on a grid and their visible neighbours were all other household agents within a specified number of cells in each of the four cardinal directions (up, down, left and right). The extent to which each household might be influenced by policy authorities' exhortations, by other households or simply by their own preferences were set at random. Among their visible neighbours, each household agent came to value most the examples set by neighbours whose water use patterns were closest to their own.

4.1 Results and Stakeholder Involvement

The results obtained from the first version of the prototype confirmed that combinations of agents, some of whom are most influenced by authority and some by word of mouth communication, can be effective in managing demand. The extent to which each agent was influenced by authority, neighbours or not at all was determined at random so that each agent was unique. A typical pattern of drought and demand is depicted in Fig. 2 which shows that consumption fell during months when the authorities were exhorting households to restrict consumption and recovered immediately the exhortation ended. This pattern is seen even more clearly in Fig.3.

This model and simulation results were presented to stakeholders from the relevant government department, regulatory authorities and water supply companies at a meeting in Oxford early in 2000. Some of the stakeholders pointed out that the simulated consumption patterns differ from what they have observed in that normal levels of water use recover more slowly than indicated by the model. The particulars of the representation of cognition were therefore changed in the revised model to accommodate that observation.

In general terms, the Thames prototype model was developed to assess the creation and strengthening of sources of information and the values placed on each information source by the agents representing domestic water users. The modelling of the influence of information sources closely followed the consistency principle. As reported below, this approach took us a considerable way towards a model of domestic water use that was found plausible by key stakeholders from the water supply industry and the relevant government authorities. But it did not take us all of the way. As a result of discussion with the stakeholders, the model was revised to replicate better their observations of water use patterns over the course and in the aftermath of recent droughts in the United Kingdom. The modifications to the model not only improved the plausibility of the output but also provides a natural means of extending the model to take account of technological changes in domestic water-using activities.

Water consumption by the domestic household agent in the model is determined by a rule specifying the environmental conditions in which a particular frequency and use-per-event would be adopted. In the first version of the model, exhortation by the policy agent took the form of the broadcast of the conditions and actions forming a water consumption rule. The conditions were two aspects of the state of the environment: temperature and dryness. The temperature aspect could take on of the values in {*warm*, *cold*} and the dryness aspect could take one of the values in {*wet*, *dry*}. The environment was *dry* when soil water was less than 85 per cent of capacity for at least two consecutive months. This was calculated from a hydrological model integrated with the social, agent-based model. The {*warm*, *cold*} aspect had no effect and, so, was set randomly. The water-use rules thus had the form:

if (environmentalState [warm dry])
then (activityFrequencyConsumption ?activity ?freq ?cons)

Household agents in the first model version would remember both the *if* part of the rule and the *then* part of the rule and would attach to the rule an endorsement containing the source of the rule (policy authority, observed behaviour of a neighbour or self sourced). This was the main reason why domestic consumption reverted immediately to normal when exhortation ended. The reason for the end of exhortation was that soil water increased to at least 85 per cent of capacity and so the dryness aspect of the environment changed from *dry* to *wet*. The rules that were adopted during the drought simply ceased to be relevant until the next drought. The rules themselves remained in memory, however, so that at the next onset of dryness, water demand declined immediately.

Consequently, the first change made to the representation of cognition in the revised model was to change the form of the exhortation from an *if-then* rule to the *then* part of the rule – the consequent – alone. As part of the same change, the household agents were respecified to note, remember and evaluate the importance to them of the consumption frequency and use-per-event of their visible neighbours rather than neighbours' consumption *and* prevailing environmental conditions. This was an essential revision to enable the drought-induced consumption patterns to persist in the aftermath of the drought.

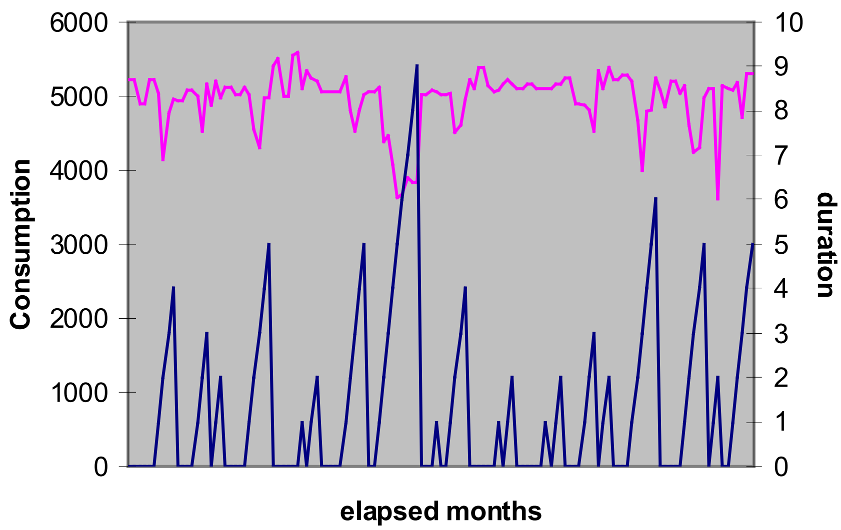


Fig. 2. The lower saw-toothed line is the number of consecutive months of drought and the upper line is the simulated domestic water use. By inspection, the volume of water consumption declined during the dry periods which was when the policy agent was exhorting households to restrict their consumption and recovered as soon as exhortation ended

As a result of this respecification of agent cognition and communication, the effect of drought and exhortation was a secular decline in water use. That is, each exhortation resulted in a reduction in water use that was not subsequently reversed. This was clearly too effective a change since water consumption does revert to normal after a drought and period of voluntary or statutory restriction even if it does so more gradually than in the first model.

The reversion to normal consumption was obtained by defining a normal level of consumption for each agent in each water using activity and also by allowing other interesting aspects of the action – mainly its source – to be forgotten over time. The consumption norms were generated at random for each household agent at the time of its creation. The norms for frequency and use-per-event were endorsed as such by each agent and were remembered throughout the simulations. All other aspects of a consumption action would be remembered with a probability that is directly related to its importance to the agent and inversely related to the lapse of time since the endorsement was attached to its object. This mechanism, suggested by Anderson's [2] work on memory and recall, sets the probability of remembering a feature of an action as vt^{-d} where v is a measure of the importance of the memory to the agent, t is the number of months since the token was attached to the action or agent and $d>0$ is the factor determining the rate of decay of the probability of remembering the endorsement. The value of d was determined at random for each agent with a maximum set by the model operator for each experiment.

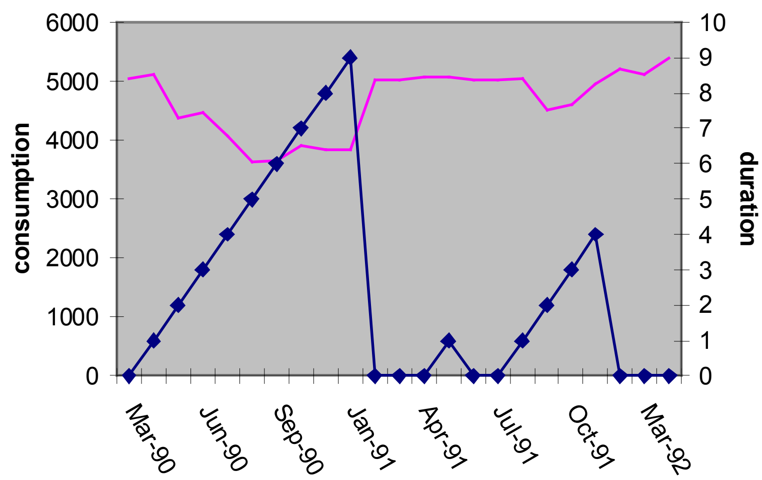


Fig. 3. A representative segment of the time series reported in Fig. 2, using the soil water data simulated for the period from March, 1990 to April, 1992. The dip in consumption from April, 1990 until January, 1991 corresponds exactly to the period of exhortation by the policy agent. There are no lags and no residual consumption dip after the end of the dry period and policy agent's exhortation

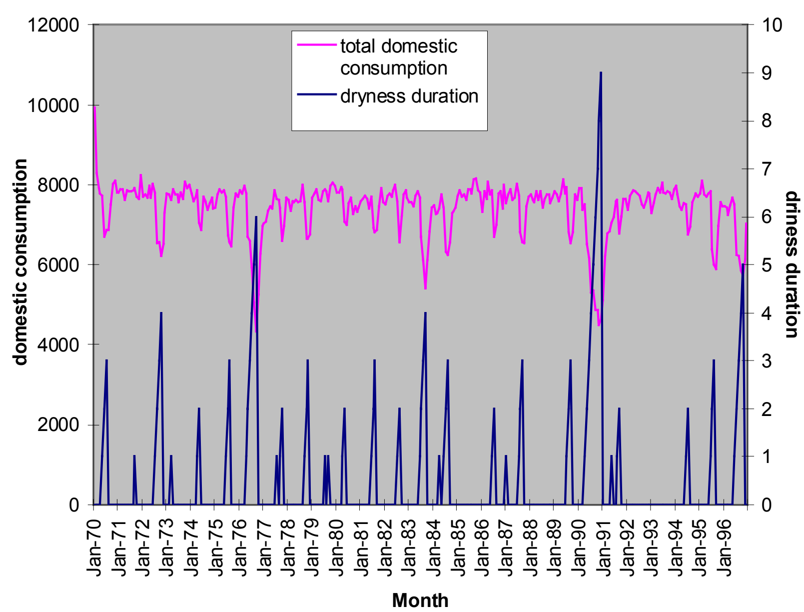


Fig. 4. Throughout the period from 1970 to 1997, consumption recovered from exhortation induced reductions over a period of months after the drought spikes in the chart indicate and end to drought and therefore exhortation

The effect of these changes on the simulated response of water use to exhortation, as seen in Figs. 4 and 5, was create the lags in consumption observed and expected by the stakeholders.

4.2 Implications for ABSS and IAM

It is clear in the first place that the initial model and the revisions thereto were intended to be descriptive with the description validated by the stakeholders and domain experts. Further validation is to be undertaken on the basis of more fine grained data and by the identification of actual water using activities and the normal frequencies and water consumption per use-event. In this way, the effects of (for example) technological changes in water using activities – perhaps toilets, washing machines or dishwashers using less water – can be modelled directly at finer grain and represented in more coarse grained models by changes in water use associated with a smaller number of abstract activities. As a result, it will be possible to model a household sector in a model of a whole region without detailed representations of the households. At the same time, whenever stakeholders want to know how the simulated technological changes influence household agent in detail – perhaps to assess the plausibility of the coarse grained models – they will be able to investigate that behaviour. Provided that the models are consistent across grains of analysis, the whole suite of models will achieve the desiderata identified in section 3.

The importance for the IAM community of this model and its extensions is that it precludes none of the properties claimed for agent based social simulation and exhibits several of the important properties such as the ability to interrogate the model and the consequent transparency of the modelled processes as well as the validation by stakeholders and in relation to observed historical outcomes.

5 Discussion: Modelling Climate Policy Using Agents

Collaboration between the ABSS and the IAM communities holds substantial advantages for both. We have argued at length and sought to demonstrate with a prototype model that representational ABSS methods combined with agent oriented software engineering methods meet the most stringent criteria of good integrated assessment modelling better than any extant alternative. The benefits for the ABSS and, more generally, the MAS community is the focused development of techniques for specifying and analysing huge, complex systems. Social systems are a source of analogy for mechanism design in MAS – especially for electronic commerce and the interrogation of large federated databases and the like.

The validation of agent based integrated assessment models requires the participation of domain experts, usually stakeholders, to assess the plausibility of the models. These models must therefore be good, accurate (though perhaps coarse grained) descriptions of the natural and social systems of interest and the integrated models of these systems must yield interactions with plausible statistical signatures and descriptions of more qualitative phenomena. Consequently, developing agent based simulation models for the analysis of climate change will stretch our descriptive capacities and put a premium on the ability to build a suite of consistent models supporting a process of compositional validation. The development of these technical capacities will create a body of social simulation technique supporting interrogation of the simulation system both to validate the models and to make them comprehensible by stakeholders and other users.

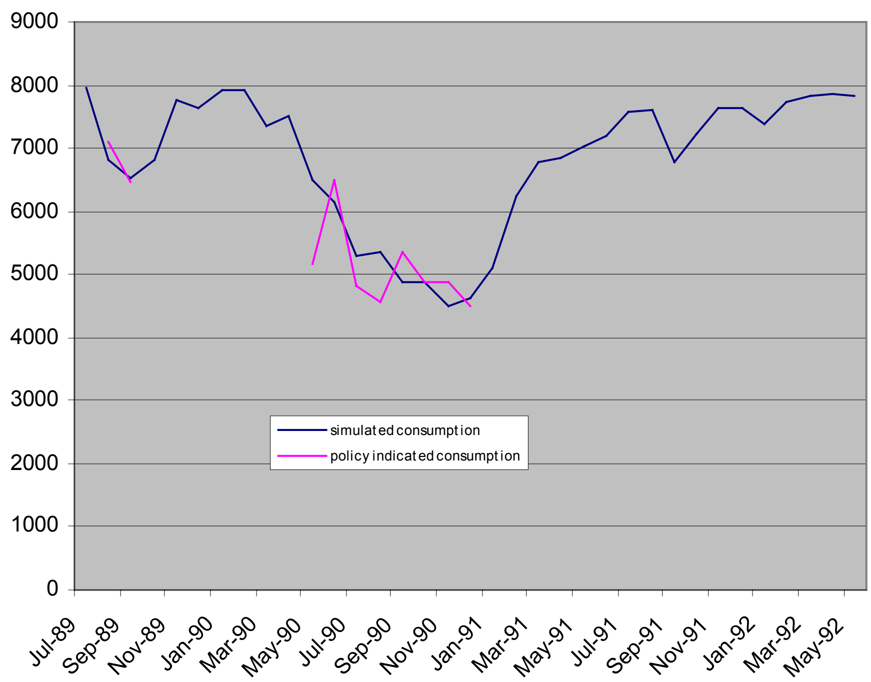


Fig. 5. The relationship of consumption to exhortation in the simulation experiments covers a relative short drought in the summer of 1989 and the longer 1990-91 drought. In both cases, simulated domestic users reduced consumption in line with exhortation. An interesting result, requiring empirical and expert confirmation, is that the reduction in consumption was longer lasting and the recovery more gradual after of the longer period of exhortation and deeper cuts in water use in 1990-91 in comparison with the relatively brief drought and period of exhortation in 1989

As with many good innovations, developing the solution to an applied problem in one area provides, or leads directly to, the capacity to solve a general class of problems. The problems to be solved in the application of agent based simulation techniques to integrated assessment modelling require the development of solutions to problems that appear widely in social analysis and software engineering for multi agent systems.

Acknowledgements

All of the co-authors of this paper are participants in the EU FIRMA project (http://www.cpm.mmu.ac.uk/firma/) and the paper benefits from discussions with colleagues in that project. The model reported here was produced as part of the FIRMA project (EVK1-CT1999-00016).

References

1. Alcamo, J., Leemans, R. and Kreileman, E., eds. Global change scenarios of the 21st Century: Results from the IMAGE 2.1 model. Pergamon, Oxford, 1998.
2. Anderson, J.R. Rules of the Mind, Lawrence Erlbaum Associates, Hillside, NJ, 1993.
3. Brown, R., Social Psychology, Collier Macmillan, London, 1965.
4. Byrne D. and G.L. Clore and G. Smeaton, The Attraction Hypothesis: Do Similar Attitudes Affect Anything?, Journal of Personality and Social Psychology, 1986
5. Edmonds, B., (2000)
6. Grubb, M. et al.. Energy Policies and the Greenhouse Effect, Volume 1: Policy Appraisal, RIIA/Dartmouth, 1990.
7. IPCC. IPCC Second Assessment Report: Climate Change 1995. Intergovernmental Panel on Climate Change, 1996.
8. Jasanoff, S., and Wynne, B.. Science and decision making. In: Rayner, S. and Malone, E.L., Eds. Human Choices and Climate Change. Volume 1: The Societal Framework. Batelle Press, Columbus, Ohio, 1998.
9. Moss, S., B. Edmonds, and H. Gaylard, "Modelling R&D Strategy as a Network Search Problem", Centre for Policy Modelling Technical Report 96-11, 1996.
10. Moss (1999)
11. Nordhaus, W. Managing the Global Commons. The Economics of Climate Change, MIT Press, Cambridge, MA, 1994.
12. Pahl-Wostl, C. Integrated Assessment of Regional Climate Change and a new Role for Computer Models at the Interface between Science and Society, in: Prospects for Integrated Environmental Assessment: Lessons learnt from the case of climate change A. Sors, A. Liberatore, S. Funtowicz, J.C. Hourcade, J.L. Fellous, European Commission, DG XII, Toulouse, pp. 156-160, 1996.
13. Paterson, M. Global Warming and Global Politics. Routledge, London. 1996.
14. Rayner, S. and Malone, E.L.. Human Choices and Climate Change. Volume 1: The Societal Framework. Batelle Press, Columbus, Ohio, 1998.
15. Rotmans, J.. Methods for IA: the challenges and opportunities ahead. Environmental Modeling and Assessment 3(3): 155-179, 1998.
16. Rotmans, J and de Vries, H.J.M. Different perspectives on global futures: the TARGETS approach. Cambridge University Press, Cambridge, 1997.
17. Rotmans, J., H. Dowlatabadi and T. Parson Integrated Assessment of Climate Change: Evaluation of methods and strategies, in: Human Choice and Climate Change eds S. Rayner and L. Malone, Battelle Press, Columbus Ohio, 1998.
18. Schellnhuber, H.J. 'Earth system' analysis and the second Copernican revolution. Nature 402 (6761) , 1999.
19. Skolnikoff, E.B.. The Elusive Transformation: Science, Technology, and the Evolution of International Politics. Princeton University Press, 1994.
20. Thompson, M., Ellis, R. and Wildavsky, ACultural theory. Westview, Boulder, 1990.

Enhancing Multi-Agent Based Simulation with Human-Like Decision Making Strategies

Emma Norling[1], Liz Sonenberg[2], and Ralph Rönnquist[3]

[1] Computer Science and Software Engineering
The University of Melbourne
[2] Department of Information Systems
The University of Melbourne
[3] Agent Oriented Software Pty Ltd

Abstract. We are exploring the enhancement of models of agent behaviour with more "human-like" decision making strategies than are presently available. Our motivation is to build multi-agent based simulations of human societies that would exhibit more realistic simulation behaviours. This in turn will allow researchers to study more complex issues regarding individual and organisational performance than is possible with present systems. Drawing on studies into naturalistic decision making, which looks at people's decision making in their natural environments, we propose ways of integrating a recognition-primed decision making approach with the popular BDI agent architecture, and report on preliminary empirical studies.

1 Introduction

Multi-agent based simulation provides a powerful tool for organisations such as the military and emergency services to evaluate equipment and procedures before their use in operational situations. Modelling physical systems, such as aircraft, fire hoses and radar, is relatively straightforward, given their specifications. Modelling the human operators in the system is a far more difficult problem.

There are already multi-agent simulation systems of this sort in operation, such as Tac-Air-Soar and SWARMM [11,21]. Systems like these are useful tools as they are, but with more accurate models of human cognition, they could support more detailed analysis of individual and organisational performance.

Here we focus on the way agents choose what course of action to perform. Many agent architectures, at least theoretically, use rational choice for decision making, whereby all possible choices are "scored" and the highest ranked is chosen. While this should lead to the optimal choice being made, evidence suggests that people rarely use this type of decision making in their natural environments [17]. The field of psychology called "naturalistic decision making" (NDM) offers several different descriptive models of decision making which are applicable in different sorts of situations, which we will investigate for agent models.

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 214–228, 2000.

Several candidate agent architectures are grounded in models of human reasoning, such as ACT-R [2], BDI [19], COGNET [5]), and Soar [15]. We have chosen the Belief-Desire-Intention (BDI) agent model for our work, partly because some of its features suggest that it will integrate well with the forms of decision making we seek to explore (see below), and partly because of its value in an existing multi-agent simulation, SWARMM (see section 1.1) which can provide a potential test-bed for our work.

We see the BDI framework as a useful starting point for modelling human operators. It provides goal-directed behaviour, whereby an agent's actions are motivated by a hierarchy of goals rather than being purely reactive. The notion of intentions is also useful – having an intention to perform an action (or course of actions) implies some commitment to these actions. This prevents complicated reasoning at every time step, since once an agent has decided to do something, it will continue to do it until it becomes either impossible or unnecessary (when the goal it set out to achieve is no longer pertinent). Another feature of the BDI architecture is the structure of knowledge within an agent – beliefs, goals and plans – which in our experience can simplify communication between the programmers and the domain experts.

Integrating NDM into the BDI architecture can provide useful enhancements to existing systems, as well as increased scope for future development. Furthermore, we hope that lessons learned in integrating NDM into the BDI architecture will also be applicable to other architectures.

The remainder of this paper is organised as follows: First we give some examples of the sorts of MAS we are interested in, outlining the MAS in operation at the Australian Defence Science and Technology Organisation, Air Operations Division, and various other application areas. Next, we describe naturalistic decision making, which studies how people operate in their natural settings. Thirdly we look at the BDI agent architecture and discuss how we might implement a NDM model within this framework. We conclude with a discussion of our preliminary results and future work.

1.1 Smart Whole AiR Mission Model (SWARMM)

SWARMM was developed by Air Operations Division of the Australian Defence and Science Technology Organisation, in conjunction with the Australian Artificial Intelligence Institute (AAII). It is used to simulate fighter aircraft operations for the Royal Australian Air Force (RAAF). Each pilot in the system is an agent, programmed with dMARS, a BDI-based programming environment developed at AAII [6]. The physical models used in the simulation are implemented in FORTRAN and C. The agents receive data from the physical models equivalent to the information a real pilot would receive from his/her vision and instruments.

Each agent in the simulation goes through a cycle of:

1. *Situation awareness*, when the raw data from physical models is transformed into more complex symbolic descriptors;

2. *Situation assessment*, when the situation is characterised based on the output of the previous step;
3. *Tactics selection*, when a tactic is selected based on the current set of goals and the situation that was recognised in the previous step; and
4. *Operating procedures*, when a standard operating procedure is selected to implement the chosen tactic

The SWARMM programmers have had more than a decade of close contact with fighter squadrons, through interviews, mission briefing and involvement in training exercises. This experience has confirmed the value of the above cycle. It allows simple integration of documented standard procedures, which are recored in a similar fashion, and expert pilots can fairly comfortably describe their knowledge in this form. This makes the transfer of knowledge from pilot to programmer and feedback from programmer to pilot easier, since they then talk in similar terms. Importantly it also means that RAAF personnel are confident in the results of these simulations, due to their involvement in and understanding of the development process. [10]

SWARMM models squadrons of fighter pilots, with a heavy emphasis on teamwork. It is used to test new equipment and tactics, and has proved extremely useful for this purpose. However, one of the current aims of the project is to use these same agents in human-in-the-loop simulations. It is feared that while the teams of pilots do behave like humans to the outside observer, a human flying in a simulator either against or in teams with them would not find them to be particularly believable. The ideas explored in this paper are part of the work in this project.

1.2 Other Applications

A good model of human behaviour has applications in many different types of systems, not just military simulation. As mentioned above, simulations similar to those used by the military are used in other areas, such as disaster recovery, where it is difficult and/or costly to stage a real-life rehearsal. Another key application area is in the computer games industry – many games companies these days use their "sophisticated AI" as a key marketing strategy. However a common complaint from games players is that while the AI in games is improving, playing against a computer generated character does not compare to playing against another human player. Work in this area is being explored by researchers such as John Laird [14].

2 Naturalistic Decision Making

> "The study of NDM asks how experienced people, working as individuals or groups in dynamic, uncertain, and often fast-paced environments, identify and assess their situation, make decisions and take actions whose consequences are meaningful to them and to the large organisation in which they operate." [22]

"Naturalistic decision making" is a term that first emerged at a 1989 workshop organised for researchers who were studying decision making in realistic settings (e.g. medicine, nuclear power plants and executive planning). Their studies were showing that classical decision making theories (i.e. those based around rational choice) simply did not apply to these real world settings. A number of factors made it difficult or impossible to apply rational choice theory, and evidence showed that even when they were trained to use this strategy, decision makers rarely did ([12], p. 99).

What has emerged from the study of NDM over the past ten years is a better understanding of the way we make decisions in complex situations. In some cases we do apply rational choice theory, where several options are generated and the "best" selected, but other strategies are far more commonly used. The research in this field is primarily aimed at designing decision aids and training that complement expert decision making styles, but we believe the findings can also be used to develop better cognitive models for the humans in our simulations.

Orasanu and Connolly [18] list eight factors which they claim characterise naturalistic settings. Many classical decision making studies ignore or deliberately limit these factors, which makes rational choice theory easier to apply. These factors are:

- Ill-structured problems
- Uncertain dynamic environments
- Shifting, ill-defined, or competing goals
- Action/feedback loops
- Time stress
- High stakes
- Multiple players
- Organisational goals and norms

Not all factors will be present in every naturalistic setting, but each adds complexity to the problem.

2.1 Models of Naturalistic Decision Making

Several models of naturalistic decision making have been proposed, but as yet, no single one of them captures the full range of decision making in naturalistic settings. Lipshitz [16] gives a summary of nine models, which he points out are not contradictory, but illustrate the different types of decision making that can be used.

Note that researchers in this field are generally interested in people who are experienced in their field. Hubert Dreyfus explains "the five-stage model of skill acquisition" in [7], with levels ranging from novice (someone just starting out in the field, such as a learner driver), through to expert (someone who is immersed in the world of his/her skillful activity). Most NDM models assume some level of expertise in the field, not necessarily expert, but definitely not novice.

One of the best known models is Klein's recognition-primed decision making (RPD) model [12], illustrated in Fig. 1. As Klein himself points out, "The RPD model is not synonymous with NDM research," ([12], p.102), but studies of experts in various fields show that a large proportion of their decisions are made this way (see table 7.2 in [12]). The important thing to note about this model is that the focus is on situation assessment. Once the decision maker has recognised the situation, there are four by-products:

– he/she expects certain things to occur but not others;
– he/she pays attention to certain cues to support the diagnosis;
– there is some understanding of what goals are plausible to achieve;
– and there are certain actions which are likely to succeed.

The decision maker then selects an action, runs a quick mental simulation of the action, and if he/she thinks it will succeed, implements it. Once the course of action has been selected, the the situation is monitored to make sure it remains as expected, and if not, alternate actions may need to be considered, but otherwise the decision maker does not consider other options. (Note that this aspect of the model is not indicated in the diagram.) In the RPD model, the experienced operator will usually select a course of action "automatically" once the situation is recognised, and that course of actions is likely to be one that was previously successful in that situation.

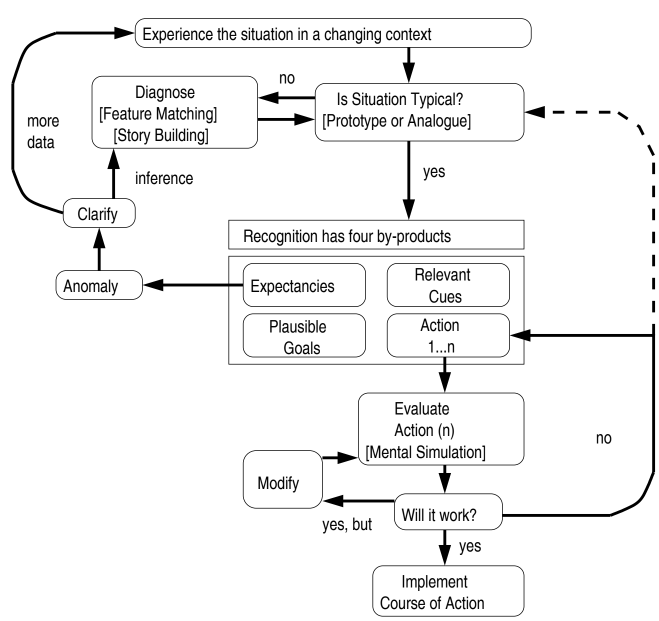


Fig. 1. Klein's integrated version of recognition-primed decision model (Figure 7.1 from "Sources of Power" by G. Klein 1998, [12])

This model captures the idea that as a person gains expertise, they are better able to recognise subtle differences between situations, and choose appropriate courses of actions accordingly.

3 Belief-Desire-Intention (BDI) Agents

BDI agents are based on the philosophical concepts of intentions, plans and practical reasoning developed by Bratman [3]. This model is based in folk psychology, that is, the way that we *think* we think. The model provides a useful first approximation to human cognition, but there is much scope for refinement.

The *beliefs* of an agent are its view of the world, which is not necessarily the same as the state of the world, because the sensors may be imperfect. The information supplied can be both incomplete and noisy.

Rather than the *desires* of an agent we refer to its goals. These give the state of the world in which the agent wishes to be, and must be consistent.

Its *intentions* are the plans that it is currently executing. There may be more than one current plan, because an agent may be simultaneously working towards multiple (non-conflicting) goals. Once an agent forms an intention (i.e. selects a plan) it is in some sense committed to that plan – it will continue executing it (or at least have an intention to execute it) until the goal is achieved, the goal becomes irrelevant, or it is impossible to proceed with that plan.

A *plan* is a "recipe" to achieve a particular goal. It is a sequence of actions and/or sub-goals to achieve. If any step in the sequence fails, the plan itself will fail. One of the features of a BDI system is that when a plan fails, the agent will recover (if possible). It will try to find another way of achieving the goal, taking into account the fact that the world (and hence the agent's beliefs) is changing. An agent stores its plans in a plan library.

The agent shown in Fig. 2 goes through a continuous cycle of:

1. sensing the environment
2. reasoning about beliefs, goals and intentions
3. performing one or more actions

This cycle is very similar to the one used in SWARMM (section 1.1), which separates step 2 into two stages: situation assessment followed by tactics selection. Indeed, SWARMM is implemented using a BDI architecture.

During the reasoning stage of the cycle, the agent must reason about beliefs (if and how they should change), goals (changes in beliefs may affect feasibility of goals), and intentions (changes in goals may cause the agent to drop some intentions and/or form new ones). The agent must also decide which action (or actions) to perform next, from the current intentions.

When there are multiple plans available to achieve a given goal, the agent, in theory, uses rational choice to select a plan [4]. That is, the merits of all the applicable plans are evaluated, and the "best" one is selected. However NDM

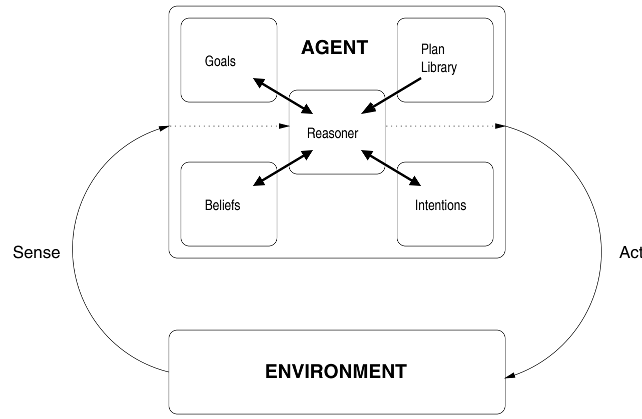


Fig. 2. The structure of a BDI agent

research indicates that this is not the way that people make decisions, and this is where we are currently working to improve the BDI model. [1].

In practical implementations of BDI architectures, such as JACK [1] or dMARS [6], each plan is designed to handle a particular goal in a particular context. (In these systems, 'context' is a set of conditions which the agent must believe to be true.) This allows the programmer to specify different ways of achieving the same goal in different situations, but it is also possible to have several plans applicable in a given situation (when the contexts overlap).

3.1 An Example of Decision Making in a BDI Framework

Example 1. Suppose an agent has a goal of getting from one side of the business district of town to the other (a distance of about 3km). She may have three plans for handling this goal: the first involves walking; the second involves driving; and the third involves catching the train. Each of these plans has a context. The walking plan might specify that it is only applicable if the agent has at least 40 minutes to achieve the goal, and that it is not applicable if it is raining heavily. The driving plan might indicate it is only applicable if the car has fuel. And the context of the train plan might be that it is only applicable between 6am and midnight (trains don't run all night).

This example is a simple one, and doesn't have many of the features that characterise naturalistic settings, but is useful for our preliminary work. One can see that there are three distinct cases in the example:

[1] In practical implementations of BDI systems, such as JACK or dMARS, plan selection strategies are more varied. The programmer can implement meta-level reasoning for deliberation, or rely on the default strategies of the programming language, which place an ordering on the plans.

1. when there is only one plan that is applicable (e.g. it is 1am and raining heavily);
2. when there is no plan applicable (e.g. it is 1am, raining heavily and the car is out of fuel); and
3. when there are multiple plans applicable (e.g. it is 1pm, the weather is fine, the car has fuel, and the agent has 40 minutes available for the trip).

If there is only one plan applicable, the agent simply forms an intention to execute that plan. If there is no applicable plan, the goal will fail – the agent may have a higher level plan to deal with this, or not, but that is not discussed here. When there is more than one applicable plan, the agent must select one of them.

4 Implementing Recognition-Primed Decision Making in a BDI Framework

Recognition-primed decision making and the BDI model of agents already have some overlaps. Both have goal-directed behaviour, and both employ the notion of commitment to a course of action once it has been selected. What we hope to achieve by integrating RPD into the BDI framework is a more naturalistic way of tackling the problem of which plan should be chosen when there are multiple applicable plans. Readers should note that there is a separate issue of what to do when there are no plans applicable, but that RPD does not attempt to address that problem, and it is not considered here.

There are several ways in which recognition-primed decision making could be used with BDI agents. Here we describe three possible approaches, with increasing complexity and scope. These approaches are related to work in other areas of artificial intelligence.

4.1 A Naive Approach

A naive approach to using RPD would be to assume that the agents are experts, and hence able to recognise all possible situations. It would be up to the agent designer to ensure that all possible scenarios were considered, with a plan (and only one plan) for each one. In this case the agent would never have to make a choice between plans, so the issue of a choice strategy is totally avoided.

This approach has similarities to case-based reasoning [13], but has several problems. The first is the assumption that the agents are experts. For some cases this may be a valid assumption, but it severely limits the number of cases that can be handled. A related problem is the assumption that experts will be able to distinguish between and recognise *all* situations. While studies do report a very high proportion of recognised situations (e.g. on average 90% of decisions by oil rig managers [9]), no expert learns to recognise all situations in the sorts of complex domains we would like to consider. Yet another problem lies in expecting the designer to anticipate every possible situation and provide a plan for each one – once again, the complexity of the domain can make this infeasible.

4.2 A Preference-Based Approach

A slightly more complex approach is to have a preference weighting on plans. Initially all plans would be equally weighted. When an agent was faced with a choice between several applicable plans, it would pick the one with the highest preference weighting (or randomly choose from those with the highest preference weighting). If the plan succeeded, the preference weighting would increase, and if it failed it would decrease.

Using Ex. 1, say it is 8:20am, the weather is fine, and the agent needs to be at a meeting at 9am. All three plans are applicable and all have equal preference weighting. The driving plan is randomly selected. However, the agent gets stuck in traffic and this leads to her being late for the meeting. As a result, the preference weighting on this plan is decreased, so she will be less likely to choose driving next time.

This approach is a simple form of reinforcement learning [20]. It captures the idea that we are more likely to consider something that worked before, but it fails to capture the idea that experts learn the subtle differences between situations.

4.3 A Context-Based Approach

The third approach is to use a form of learning to adjust plan context. The agent would need to record the world state each time a plan is used, and then use reflection to work out what caused the failures. This could then be used to refine the context of a plan, a process that should eventually lead to there being no overlaps (or cases where plans are equally good, in which case choice could be random).

Again using Ex. 1, let's say it is 9am, the weather is fine, the car is fully fuelled, and the agent wants to get to the other side of town by 10am. In this case, all three plans are applicable (she could walk, drive or catch the train). So she randomly chooses a plan, and decides to drive. However, reaching the other side of town, she finds that there is nowhere to park the car, and so the plan fails. She knows what part of the plan failed, but without having other examples of using this plan, it is difficult to know what caused the failure. With enough examples, she may conclude that between 8:30am and 5pm she will not be able to get a car park, so should walk or catch the train.

Note that an agent, just like a person, will never be able to anticipate all of the things that will cause a plan to fail. For example, if she tried to catch the train, but someone had fallen on the tracks and this had stopped the train, this would cause the plan to fail. However, a context saying "if no one has fallen on the tracks" would not be particularly useful. For one, this event would likely occur after the agent had commenced using the plan. And secondly, it is not something that is likely to occur often enough (we hope!) for us to want to capture it – otherwise we would end up with a myriad of plans for incredibly specific situations.

In this approach, the idea of refining one's ability to distinguish between situations is captured. In the cases where there are multiple equally "good" plans (i.e. that are all likely to succeed), a preference-based approach should be added, so that the idea of preferring to use plans that have worked before is also captured.

4.4 Implications for Implementation with BDI Agents

Of these three approaches, the first needs no modification of the existing BDI architecture – it is up to the programmer to implement the agents plans as described. For a simple scenario such as the one we have described, this is an easy task. The number of influencing factors are limited, and the programmer knows exactly what they are. However as the simulation environment becomes more complex, and the agent programmer does not know the exact details of its implementation, this task becomes increasingly complex.

To implement the preference-based approach, we take advantage of the meta-level reasoning capabilities of JACK. We can keep a record of which plans have failed and succeeded, and calculate a rank for each plan from this record. In the simplest case, we simply increment the rank on success and decrement it on failure. To simulate a "fading memory", the weight of successes and failures decrease over time.

The difficulty in implementing the context-based approach is that BDI agents generally use a fixed plan library. An agent's beliefs change at runtime, but the plans themselves, including the context conditions, do not change. Work on allowing changing context in JACK will shortly commence. There are two possible ways of approaching context change, as discussed in [8]. One can take a "top-down" approach, where the use tries to lay down general a-priori principles, or a "bottom-up" approach, where context is deduced through experimentation. Ultimately, one would like an agent to deduce the appropriate context from its experience, but this leads to an important question: How does the agent recognise contextual differences *at the appropriate level of abstraction*? In the example given above, where no car park was available, it would be silly to change the context of the driving plan to say "don't use this plan if it is 9am and the weather is fine." However an appropriate modification to the context may be: "don't use this plan if it is after the start of morning rush hour and before the start of evening rush hour." Obviously the agent would need more than a single failure to be able to make this deduction, but exactly how much experience would it need? When should it decide to modify the context? It may be more practical to give "hints" to the agent – for example, knowledge that the morning rush hour is from 7am-9am, and the evening from 4:30pm-7pm.

5 Preliminary Results and Future Work

A simulation of the world described in the example above has been constructed. It communicates with agents via sockets using a simple protocol of atomic actions. Agents may move one "step" (either walking or driving), check the weather, wait for a train, catch a train, or park the car. The world sends a message to the agent when the action has been performed, and an agent may send only one action request at a time. Traffic speeds (and hence travel times), train frequencies and time to find a park are dependent on the time of day.

The preference based-model has been explored with several variations on the method for calculating rank. Figure 3 shows one of these variations. In this case, we have a risk-averse agent, which penalises failure at twice the rate that it rewards success. It also "forgets" successes and failures as time goes by. The graph shows how the rank of the three plans (walking, driving and catching the train) changes over a number of attempts. In the first 15 attempts, the agent has to get across town during the morning rush hour, and in the following 6 attempts it has to make the trip in the evening. The weather is fixed at fine, and the agent has 40 minutes for each trip, so all three plans are applicable in both cases.

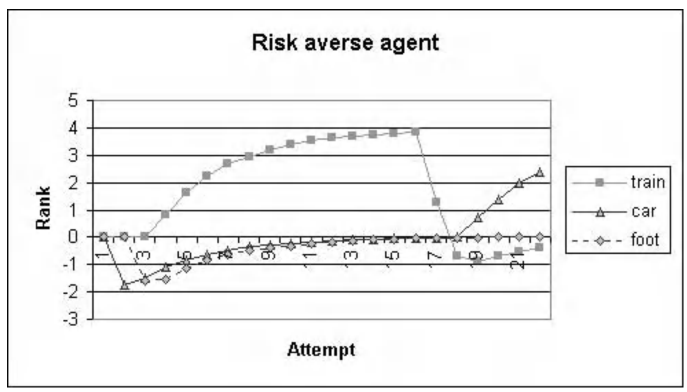


Fig. 3. Changes in rank over time with the preference-based approach

The rank is initially zero for all plans, and is calculated with a numerical formula, which can be modified to increase or decrease the rate of memory loss, or change the reward or penalty scores. In the example here, the agent first tries the driving and then the walking plan, but these both fails. It then uses the train plan, which proves to be successful every time at rush hour. It then tries to catch the train when the context changes to evening (at attempt 16), but when this fails, it quickly tries a different plan. This mirrors human behaviour – if you are accustomed to doing something in a particular way, it takes less effort to keep doing it that way, even if the context changes, so long as you believe that method is still feasible.

Using this method, we can create a limited form of memory about previous success of a plan, and by manipulating the rank calculation, the agent can exhibit some human-like qualities for plan selection when the context provided by the system designer is insufficient for this choice. However the information captured is rather limited – it fails to capture *why* the plan has failed. Each time the context changes, the agent must re-learn which is a "good" plan. For some applications (particularly where context is unlikely to change often), this preference-based approach may be a close enough approximation to recognition-primed decision making, especially since there will be less computation involved than refining the context (whichever way this is done).

5.1 Future Work

The primary focus of this work is to enhance the BDI model with more "natural" decision making characteristics, without sacrificing the practical advantages of the BDI approach visible in a number of existing implemented systems.

Implementing the context-based model of recognition-primed decision making is the next stage in our work. We anticipate that this will give a better model of recognition-primed decision making than the preference-based model, but as mentioned above, we would like to explore the tradeoffs. How these tradeoffs will be compared is itself an interesting problem, which is discussed further below.

The simulation used to date has been useful in generating these preliminary results, but as mentioned above, it is extremely simplistic, and it does not display many of the characteristics of a naturalistic decision making environment. For the next stage of our work we will require a richer environment. There are already many simulation environments that are readily available, but relatively few with the richness that we require. Interestingly, one of the possible environments that does meet our requirements is a gaming environment – modelling a human player in a game such as Quake would in many ways be comparable to modelling a human in a flight simulator, without requiring any of the specialist knowledge of a pilot. Such an environment also gives the potential to study decision making in a multi-agent setting, be they human or software agents. The state and actions of others will influence the decision making of an agent, particularly when cooperation and coordination are issues.

6 Discussion

Many of the issues raised in open discussion at MABS-2000 have direct relevance to this work. In particular, the issue of validation is extremely important. Obviously the agents we are building are designed to be in some sense unpredictable, in the same way that people are. The question then arises of how we can distinguish between agents that are human-like and those that behave in some other way.

In this case, it seems that a descriptive form of validation would be most appropriate – to take data from studies of real humans and compare with the agent behaviour. The systems we are building are designed to predict outcomes, or "see what happens," so trying to predict their behaviour and validate them in this way would be counter-productive. If the behaviour of individual agents in a small-scale system can be validated descriptively, we hope that these results will be scalable.

There are a number of methods that we are considering for validation of our system:

- Human-in-the-loop testing, where a human takes the place of one of the agents in a system and evaluates whether or not the agents are behaving as another human would (and to what extent, in which areas).
- Analysis of a data trace by a behavioural expert.
- Comparison of a trace of an agent performing a task with one from a human performing the same task.

For all these methods, a difficulty lies in explaining differences – just as you wouldn't expect exactly the same trace from two different people on a complex task, you wouldn't expect the same trace from one of these agents and a human. When then is a difference a human-like variation and when is it not? Behavioural experts are perhaps in the best position to make that judgement, with their broad background in human behaviour, but experts in the task are also likely to be good judges, as they will have a wide range of experience with different people in that task.

Another issue that was raised in the workshop was the "granularity" of agents: whether agents should represent single neurons, individuals, small groups, or some other level. In our case agents are representing individuals, but a related issue is the question of the "depth" to which we model these individuals. The answer is perhaps related to the purpose of the system. When we are primarily interested in macro-level behaviour, with no focus on the individual agents in the system, a "shallow" representation may be sufficient, focusing on perhaps just a handful of pertinent variables. But if we wish to analyse micro-level behaviour, where the performance of individuals is scrutinised, we need more detail in these agents. When a human is able to participate in the simulation, interacting with the agents as if they are other people, this becomes even more important. In these cases, the agents need to be believable as people – behaving in the way a human would, without super-human abilities. For these sorts of applications, a human-like agent is extremely important, and adapting the decision making strategies of agents is one step in this direction.

Acknowledgements

Support for this work was provided by a Strategic Partners with Industry Research and Technology (SPIRT) award from the Australian Research Council and Agent Oriented Software Pty Ltd. The authors would also like to thank Clint Heinze and Simon Goss of Air Operations Division at the Defence Science and Technology Organisation (Australia), and Peter Wallis of the Department of Computer Science and Software Engineering at the University of Melbourne, for their useful discussions regarding this work.

References

1. Agent Oriented Software Pty. Ltd. JACK Intelligent Agents. URL: http://www.agent-software.com.au/jack.html. 220
2. John R. Anderson and Christian Lebiere. *The Atomic Components of Thought.* Lawrence Eribaum Associates, 1998. 215
3. Michael E. Bratman. *Intention, Plans and Practical Reason.* Harvard University Press, Cambridge, Massachusets, 1987. 219
4. Michael E. Bratman, David J. Israel, and Martha E. Pollack. Plans and resource-bounded practical reasoning. *Computational Intelligence,* 4(4):349-355, 1988. 219
5. CHI Systems Inc. COGNET - COGnition as a NEtwork of Tasks. URL: http://www.chiinc.com/cognethome.shtml. 215
6. M. d'lnverno, D. Kinny, M. Luck, and M. Wooldridge. A formal specification of dMARS. In M. P. Singh, A. Rao, and M. J. Wooldridge, editors, *Intelligent Agents IV (LNAI Volume 1365),* volume 1365 of *Springer Lecture Notes in Artificial Intelligence,* pages 155–176. Springer, 1997. 215, 220
7. Hubert L. Dreyfus. Intuitive, deliberative and calculative models of expert performance. In Caroline E. Zsambok and Gary Klein, editors, *Naturalistic Decision Making,* pages 17-28. Lawrence Eribaum Associates, 1997. 217
8. Bruce Edmonds. The pragmatic roots of context. In *Modeling and Using Context: the Proceedings of CONTEXT'99,* volume 1688 of *Springer Lecture Notes in Artificial Intelligence,* pages 119–132, Trento, Italy, 1999. Springer-Verlag. 223
9. Rhona Flin, Keith Stewart, and Georgina Slaven. Emergency decision making in the offshore oil and gas industry. *Human Factors,* 38(2):262-277,1996 221
10. Glint Heinze, Bradley Smith, and Martin Cross. Thinking quickly: Agents for modeling air warfare. In *Proceedings of the Eighth Australian Joint Conference on Artificical Intelligence,* Brisbane, Australia, 1998. 216
11. Randolph M. Jones, John E. Laird, Paul E. Nielsen, Karen J. Coulter, Patrick Kenny, and Frank V. Koss. Automated intelligent pilots for combat flight simulation. *AI Magazine,* Spring:27-41, 1999. 214
12. Gary Klein. *Sources of Power.* The MIT Press, 1998. 217, 218
13. Janet Kolodner. *Case-Based Reasoning.* Morgan Kaufmann Publishers, 1993. 221
14. John E. Laird. It knows what you're going to do: Adding anticipation to a quakebot. Technical Report SS-00-02, AAAI, March 2000. 216
15. John E. Laird, Alien Newell, and Paul S. Rosenbloom. Soar: An architecture for general intelligence. *Artificial Intelligence,* 33(1) :1–64, 1987. 215
16. Raanan Lipshitz. Converging themes in the study of decision making in realistic settings. In Gary A. Klein, Judith Orasanu, Roberta Calderwood, and Caroline E. Zsambok, editors, *Decision Making in Action: Models and Methods,* pages 103-137. Ablex Publishing Corporation, 1993. 217
17. B. A. Mellers, A. Schwartz, and A. D. J. Cooke. Judgement and decision making. *Annual Review of Psychology,* 49:447-478, 1998. 214
18. Judith Orasanu and Terry Connolly. The reinvention of decision making. In Gary A. Klein, Judith Orasanu, Roberta Calderwood, and Caroline E. Zsambok, editors, *Decision Making in Action: Models and Methods,* pages 3-20. Ablex Publishing Corporation, 1993. 217
19. Anand S. Rao and Michael P. Georgeff. Modelling rational agents within a BDI architecture. In J. Alien, R. Fikes, and E. Sandewall, editors, *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KRR-91).* Kaufmann, 1991. 215

20. Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction.* MIT Press, 1998. 222
21. Gil Tidhar, Clinton Heinze, and Mario Selvestrel. Flying together: Modelling air mission teams. *Applied Intelligence,* 8(3): 195–218,1998 214
22. Caroline E. Zsambok. Naturalistic decision making: Where are we now? In Caroline E. Zsambok and Gary Klein, editors, *Naturalistic Decision Making,* pages 3-16. Lawrence Eribaum Associates, 1997. 216

Mapping the Envelope of Social Simulation Trajectories

Oswaldo Terán[1,2], Bruce Edmonds[1] and Steve Wallis[1]

[1] Centre for Policy Modelling, Manchester Metropolitan University,
Aytoun Building, Aytoun Street, Manchester, M1 3GH, UK.
Tel. +44 161 247 6478 Fax. +44 161 247 6802
{o.teran,b.edmonds,s.wallis}@mmu.ac.uk
[2] Department of Operation Research and Centre for Simulation and Modelling,
Universidad de Los Andes. Venezuela

Abstract. In this paper we present a methodology aimed at systematically exploring the shape of the 'envelope' of simulation trajectories and the implicit theory that a simulation represents. Thus it complements methods like Monte Carlo analysis, the inspection of single scenarios and syntactical proof. We propose a method for searching for tendencies and proving their *necessity* relative to a range of parameterisations of the model and agents' choices, and to the logic of the simulation language. This will allow us to dig up conclusions about tendencies for that fragment of the simulation theory given by the explored subspace of trajectories. Additionally, we propose and exemplifies a computational procedure that helps implement this exploration by translating the MAS simulation into a constraint-based search over possible trajectories by 'compiling' the simulation rules into a more specific form.

1 Social Simulation and the Exploration of Simulation Trajectories

A social simulation necessarily abstracts from some idea about processes that produce social phenomena. Typically this means that: *firstly*, many of the simulation parameters will be in essence chosen arbitrarily and, *secondly*, that there will be indetermistic choice processes in the simulation to 'stand in' for processes which we do not want to simulate. In particular a pseudo-random number generator often 'stands in' for some aspect of a real choice made by a social actor or some unpredictable aspect of the target environment. One can think of each choice as resulting in a branch point in the simulation – where the simulation trajectory 'branches out' into a separate trajectory for each possibility. The intended content of the simulation is exactly not the individual trajectories, but the envelope of these trajectories.

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 229-243, 2000.

It may be that every branch diverges from the others so that the result is completely *contingent* upon the exact choices made. On the other hand it may be that all branches share a common tendency or converge to the others in certain aspects. This commonality could be explicitly 'forced' by the design of the simulation in the form of an explicit constraint: for example if a room has only one exit then the actors in that room may all exit by the same door eventually *whatever* the nature of their individual choice processes. In other simulations the commonality is *emergent* in the sense that it is difficult to explain the commonality between possibilities in terms of the simulation design – this is what we call emergent tendencies. This paper documents some steps in the search for ways to understand such emergent tendencies.

A simulation study can have many purposes, including these: it may help in the understanding of some phenomena and also it may inform the design of future simulations. Exploring possible simulation trajectories and analysing the resulting dynamics of the simulation are central to both these tasks. Usually there is a trade-off between the richness of the study in terms of the number of explored trajectories (sometimes related to how fine-grained the model is) and the amount of required computational resources. The finer the model the more "realistic" the simulation model will be, but also the more intricate the analysis of the simulation will be.

A typical case where this analysis is crucial is in Multi-Agent Based Social Simulation. There, modellers may attempt to generate in the lab certain "complex" behaviours in a whole population as the result of the interaction of simpler. Unforeseen behaviour of individuals and unpredictable tendencies in the behaviour of the whole population can arise [4].

The lack of alternative methodologies and tools for appropriate exploration and analysis of the dynamics of a simulation are presently a factor, which limits the comprehension of emergent tendencies. Present methods include examining individual trajectories as in Scenario Analysis [3] and statistical sampling as in Monte Carlo techniques [12]. Each of these has its limitations. It is our purpose in this paper to complement these with an alternative way of exploring and analysing the simulation by systematically and automatically mapping the envelope of *all* possible trajectories in a substantial fragment of a simulation.

2 Enveloping Tendencies in Simulation Trajectories: A Constrained Search over Possible Models

The traditional methods for examining simulation trajectories are: Scenario Analysis and Monte Carlo techniques.

Via *Scenario Analysis* trajectories are inspected one at a time and as many alternatives as possible examined. Nevertheless, it is usually unviable to map all the possibilities, as the number of alternative trajectories is far too large. Additionally, the high amount of data increases the difficult task of searching for exceptional behaviour displayed by (groups of) agents in a simulation. Moreover, it is left up to the modeller to make conclusions about the persistence and sensitivity of certain behavioural outcomes with respect to a certain range of the factors.

On the other hand, a *Monte Carlo analysis* explores the dynamics of the simulation via statistical analysis of quantitative change (or quantitative measures of qualitative changes) observed in a sample of trajectories. The sample is done over the range of possibilities given by random variables introduced in the model to simplify uncertainties. The difficulties with this are that: it tells us what is a probable outcome rather than what is a necessary outcome and it can involve the use of inappropriate statistical assumptions.

2.1 Constrained Exploration of Trajectories

We propose the use of an exhaustive constraint-based search over a range of possible trajectories in order to establish the necessity of postulated emergent tendencies. Thus a subset of the possible simulation parameterisations and agent choices are specified; the target emergent tendencies are specified in the form of negative constraints; and an automatic search over the possible trajectories performed. The tendencies are shown to be necessary with respect to the range of parameterisations and indeterministic choices by first finding a possible trajectory without the negative constraint to show the rules are consistent and then showing that all possible trajectories violate the negation of the hypothetical tendency when this is added as a further constraint. (See fig. 1).

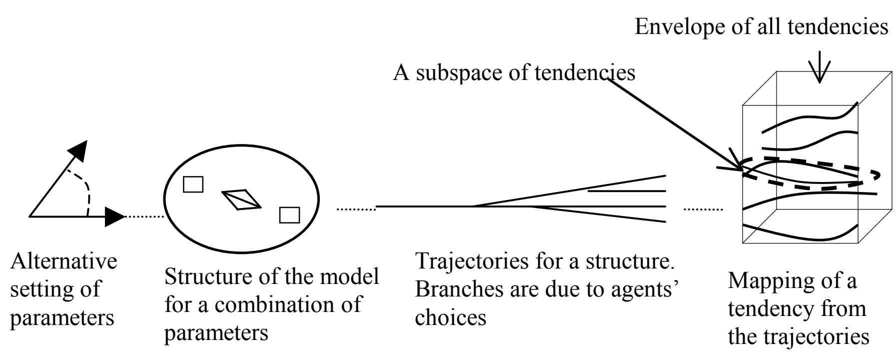


Fig. 1. A constraint-based exploration of possible simulation dynamics

2.2 Characterising the Envelope of Tendencies

In order to distinguish between the *exceptional* and the *representative* in a simulation, we will formally describe the envelope of certain tendencies in a simulation. This might be done by:

- Certain *properties* satisfied by the observed tendency.
- A *mathematical description* of a subspace of the tendencies or of a subspace given a bound of the tendencies.
- *Representative or typical instances* of such a tendency.

- A *mapping* from the setting of parameters and choices defining the trajectories to certain knowledge (maybe properties) about the tendency in such a trajectory: *(parameters* X *choices)*→*(know. of the tend.)*

2.3 Proving the Necessity of a Tendency

We want to be able to *generalise* about tendencies going from observation of individual trajectories to observation of a group of trajectories generated for certain parameters and choices. Actually, we want to know if a particular tendency is a necessary consequence of the system or a contingent one. For doing this we propose to *translate* the original MAS along with the range of parameterisations and agents' choices into a platform (described in the next section) where the alternative trajectories can be *unfolded*. Each trajectory will correspond to a possible trajectory in the original MAS. Once one trajectory is shown to satisfy the postulated tendency another set of parameters and agents' choices is selected and the new trajectory is similarly checked. If all possible trajectories are successfully *tested*, the tendency is *proved to be necessary* relative to the logic of the simulation language, the range of parameterisations and agents' choices.

The idea is to translate the MAS into a constraint-based platform in an automatic or near automatic way without changing the *meaning* of the rules that make it up in order to perform this automatic testing. In this way a user can program the system using the agent-based paradigm with all its advantages; inspect single runs of the system to gain an intuitive understanding of the system and then check the generality of this understanding for fragments of the system via this translation into a constraint-based architecture.

In the *example* shown below, all trajectories are explored for one combination of parameters, eight agents' choices per iteration and seven iterations. A simple tendency was observed characterised by a mathematical description of its boundaries. This characterisation was handled as a theorem. The theorem was proved to be necessary following a procedure similar to the one described in the previous paragraph.

2.4 What is New in this Model-Constrained Methodological Approach

It is our goal in this paper to propose an alternative approach for exploring and analysing simulation trajectories. It will allow the entire exploration and subsequent analysis of a subspace of the whole space of simulation trajectories. We are suggesting the generation of trajectories in a semantically constrained way. Constrictions will be context-dependent (over the semantics of the trajectory itself) and will be driven via the introduction of a controller or meta-module.

Like Scenario Analysis, the idea is to generate individual trajectories for different parameterisations and agents' choices but unlike Scenario Analysis the exploration is constrained to only certain range of parameters and choices.

Akin to Monte Carlo techniques it explores only part of the total range of possible trajectories. But, unlike Monte Carlo studies it explores an entire subspace of (rather than some randomly generated sample) trajectories and is able to give *definitive* answers for inquires related to the dynamics of the simulation in that subspace.

3 Towards the Implementation of a Suitable Platform for the Envelope of Trajectories: Using SDML and Declarative Programming Paradigm

SDML (Strictly Declarative Modelling Language) [8] is the declarative Multi-Agent System in which we have developed our experiments. As a source of comparisons and ideas, we have also programmed our model in a Theorem Prover [2,7,10,11]

In general, declarative programming (and in particular SDML) offers desirable features for simulation experiments as compared to imperative programming. For the social simulation community those features seem to be of particular interest when facilitating the exploring and analysis of the dynamics of the simulation [9].

3.1 Some Characteristics of Declarative Programming

- *Modularity*. Any part of the model is constructed as a group of standardised units (i.e. rules) allowing *flexibility* and *variety* in use. The declarative paradigm facilitates a greater level of modularity than the imperative paradigm because the *control* of the program is separated from the *content*. This flexibility is useful both when representing the static structure of the system and when generating the dynamics of the simulation. In our case, it facilitates the introduction of alternatives for agents' choices and parameters of the model.
- *Expressiveness*. Effective conveyance of *meaning* is a consequence of the representation of the system as linguistic clauses on a set of databases. It facilitates the interpretation of a set of social phenomena into a simulation by allowing the dual interpretation of clauses as pseudo-linguistic tokens and as entities to be computationally manipulated.
- *Easier analysis*. Context situated *analysis* of detailed data, tracks of trajectories as well as analysis of group of trajectories is much more straightforward than in imperative programs because the resulting databases can be flexibly browsed and queried.
- *The Possibility of Formal Proof.* The data generated by the dynamics of the simulation can be analysed as a *logical extension* under the particular logic of the simulation language. It opens the possibility of achieving *proofs* related to the logic of the language and the constraints imposed by the allowed choices of the agents and parameters of the model.

3.2 Other Relevant Characteristics and Features SDML Offers

- Good underlying logical properties of the system. SDML's underlying logic is close to the Strongly Grounded Autoepistemic Logic (SGAL) described by Kurt Konolige [6].
- Its backtracking procedure facilitates the exploration of alternative trajectories via the splitting of simulation paths according to agent's choices and model's parameters.
- The assumptions manager in SDML tracks the use of assumptions. Assumptions result from choices.
- A collection of useful primitives relevant to social simulation.
- The type meta-agent. A meta-agent (meta, for our purposes) is an agent "attached" to another agent as a controller; it is able to program that agent. This is used here *not* as an agent *per se* but as a module used to 'compile' rules into an efficient form as well as to monitor and control the overall search process and goals.

4 Implementing a Suitable Constraint-Based Programming Platform

The main goal of the programming strategy to be described is to increase the efficiency in terms of simulation time, thus making an efficient constraint-based search possible. The improvements will be achieved by making the rules and states more context-specific. This enables the language's inference engine to exploit more information about the logical dependencies between rules and thus increase the efficiency. Thus this can be seen as a sort of 'compilation' process, which *undoes* the agent encapsulation in order to allow the more efficient exploration of the total system behavior. In particular we split the transition rules into one per simulation period, and also by the initial parameters. This necessitates a dynamic way of building rules. This is done via a controller, which generates the rules at the beginning of the simulation.

4.1 An Efficient and Isomorphic Translation of MAS into a Single DB-RB Pair

For doing proofs efficiently we propose to carry out an isomorphic *transformation* [12] from the original MAS along with the range of parameterizations and agents' choices into a platform where the alternative trajectories can be *unfolded* more efficiently. Each trajectory will correspond to a possible trajectory in the original MAS. This transformation is a particular case of homomorphism, the two systems are equivalent, they have similar state transitions -there is a one to one translation between the states of the two systems.

This translation is needed because of the limited computational resources, the complexity of the task (usually too many trajectories have to be investigated) and the difficulties in experimenting using a MAS model directly since data and rules are encapsulated in different abstractions and hierarchies like those of agents and time levels – as a result rule dependencies are hidden. All this makes difficult and sometimes even impractical to experiment directly with models of complex systems in a MAS.

The main goal in doing such a transformation from MAS into a more efficient architecture is to take advantages of making dependencies among rules as explicit as possible and data instantiation more specific. An example where hidden dependencies are first revealed and then unwrapped exploiting them to give efficiency to the model exploration process is shown in section 5. To facilitate this we translate the original MAS into a model with a single database-rulebase pair (see fig. 2).

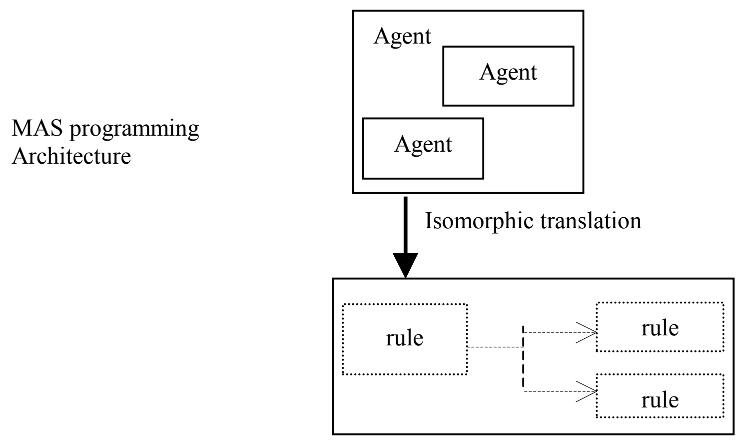


Fig. 2. Transformation of MAS system into a single rulebase-database pair

4.2 An Overview of the System

We implemented the proposed architecture in three modules; let us call them *model, prover* and *meta.* The following diagram illustrates this:

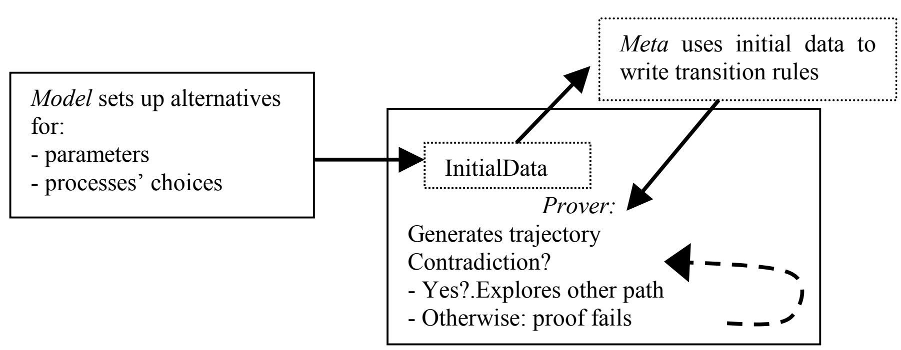


Fig. 3. Overview of the efficient implementation

4.3 Description of System Modules

We have found it convenient to *distinguish* and model as distinct entities three basic elements of a simulation: the *static structure* of the model, the *dynamics* of the simulation and the way this dynamics is "managed" by certain meta-rules or by a *controller*. Each of those entities is programmed in a different module:

1. *model, sets up the structure* of the model, that is, it gives the environment of the simulation: range of parameters, initialisations, alternative choices and basic (backward chaining) rules for calculations.
2. *prover, generates the dynamics* of the simulation. This is a sub-module of *model* (i.e. it is contained in *model*). This will basically contain the transition rules, auxiliary rules for generating pre-processing required data and the conditions to test the necessity of the theorem. All of them are rules to be executed while the simulation is going on.
3. *meta, is responsible for controlling the dynamics* of the simulation. Its meta-rules write the transition rules and the theorem in (as well as others required by) the module *prover.* A picture of the system is given in *Fig. 3.*

4.4 Program Dynamics

Modules' rules are executed in the following *sequence:*

1. *model*: initialising the environment for the proof (setting parameters, etc..)
2. *meta:* creating and placing the transition rules in *prover*.
3. *prover:* carrying on the simulation using the transition rules and backtracking while a contradiction is not found.

The program *backtracks* from a path once the conditions for the theorem are verified, then a new path with different choices and/or parameters is picked up.

4.5 Measuring the Efficiency of the Technique

Comparing the two programs, the original MAS simulation and the constraint-based translation we obtain a *speed up* by a factor of $O(NM)$, where N is the average number of agents instantiated by a rule and M is the number of simulation time steps (STI) or iterations. SDML already has facilities for discriminating among STIs, but their use is not convenient for the sort of simulation we are doing (exploring scenarios and/or proving) because of the difficulties for accessing data from any time step at any time. If we had used this facility still the simulation would have been speeded up by N. Notice that all these values are only estimations because a program stops trying to fire a rule as soon as it finds out that one of its clauses is false.

It is clear that the greater the number of entities in the simulation or the number of STIs, the larger the benefits from the technique. We must notice that the speeding up of the simulation is only one dimension of the efficiency given by the technique.

5 An Example

A simple model of a producer-consumer system, which was previously built in SDML and in the Theorem Prover OTTER, was rebuilt using the proposed modelling strategy. In the new model the exploration of possibilities is speeded up by a factor of 14. Also, the model built in OTTER, though faster than the original model in SDML, is several times slower than the improved model built in SDML.

Some of the split transition rules were the ones for creating (at each STI) producers' prices and sales, consumers' demand and orders, warehouses' level and factories' production. Among the rules for auxiliary data split were the ones for calculating: total-order and total-sales (a sum of the orders for all producers), total-order and total-sales per producer, and total-order and total-sales per consumer.

5.1 Isomorphic Transformation MAS - Single RB-DB Pair: Example of the *Rule for Prices*

The rule for prices calculates a new price for each producer at each iteration or STI (which we call *day*), according to its own price and sales, and the price and sales of a chosen producer, at the immediately previous STI.

Revealing Dependencies

Our first task, to reveal hidden dependencies, will be achieved translating the MAS into a single RB-DB pair. Each assertion in the MAS model either in the database or in the rulebase (namely each rule, predicate and function) will have an equivalent assertion in the new system. Nevertheless, the structure and dimension of predicate and function constants change to include fields referencing data implicitly in the MAS model but that have to be referenced explicitly in the new model. E.g. the agent owner of the datum and the time level the datum is valid (see fig. 4). There the rule for calculating prices at *iteration-(i+1)* for *agent P (Producer)* in the MAS has been transformed into a rule where the involved agents, *P (the rule's owner), other*P (an agent chosen by *P*) and the iteration the rule is applied (data from *iteration-i* is taken to generate data at *iteration-(i+1)*) are referenced explicitly. The transformed rule is

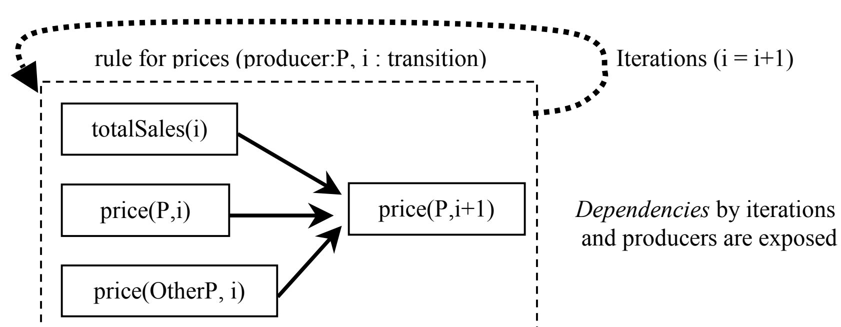


Fig. 4. Revealing dependencies, e.g. due to agent Producer "P" and time levels

shown below (notice that *day* = *i*):

```
for all (producer)
for all (consumer)
for all (day)
(
 price(producer,myprice,day)               and
 totalSales(totalSales,day)                and
 sales(producer,mySales,day)               and
 choiceAnotherProducer(anotherProducer)    and
 price(anotherProducer,otherPrice, day)    and
 calculateNewPrice(mySales,totalSales,
 otherPrice, myPrice,newPrice)
implies
 price(producer, newPrice, day + 1)
)
```

A Problem Appears: Growing of the Space of Searched Data

Nonetheless a difficulty appears after revealing dependencies: the space searched by the rules grows linearly with the number of iterations. There are too many failed attempts to fire the rule causing a serious loss of speed in the simulation over time.

Let us explain this problem continuing with the example. Note that it is intended to use the same rule at any iteration, as it was the case in the original MAS. At *iteration-1 (e.g. i = 1),* the initial data is given, then at *iteration-2* rules use data written for the single *iteration-1* and generates data for *iteration-2* (see figure 5). After this the antecedent of the rule matches data for *interation-i, i=1,2* but it only can generate new data for *i = 2.* Similarly once data for *iteration-i, i = 1, 2, …k,* has been generated the antecedent of the rule matches instances of data for all these values of *k* but it can produce new data only when *i = k.* As the simulation time goes on, the simulation becomes slower because of the discrimination the program has to carry out among the (linearly) growing amount of data matching the antecedent. In addition note that as the same rule is valid for any agent of type *Producer* the program has to discriminate among agents, this is another factor that can be improved where agents are explicitly referenced (in the example we are assuming there are three agents of type *Producer*, and consequently there will be three instances of price). In order to deal with these drawbacks we “unwrap” the rules.

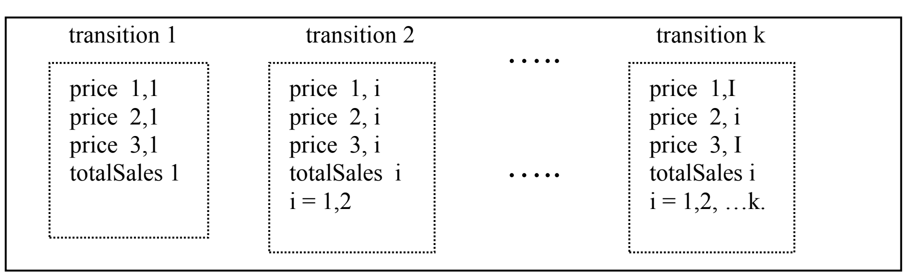


Fig. 5. Growing of the space of searched data

Dealing with this Difficulty: Unwrapping the Rules

We will split the rules for time iteration and agent. So that we will have a rule for time iteration and agent, when possible, in order to give the data to be instantiated by the antecedent of the rule as explicitly as possible (see figure 6). So, each rule will have an explicit reference to data given in previous or in the present iteration.

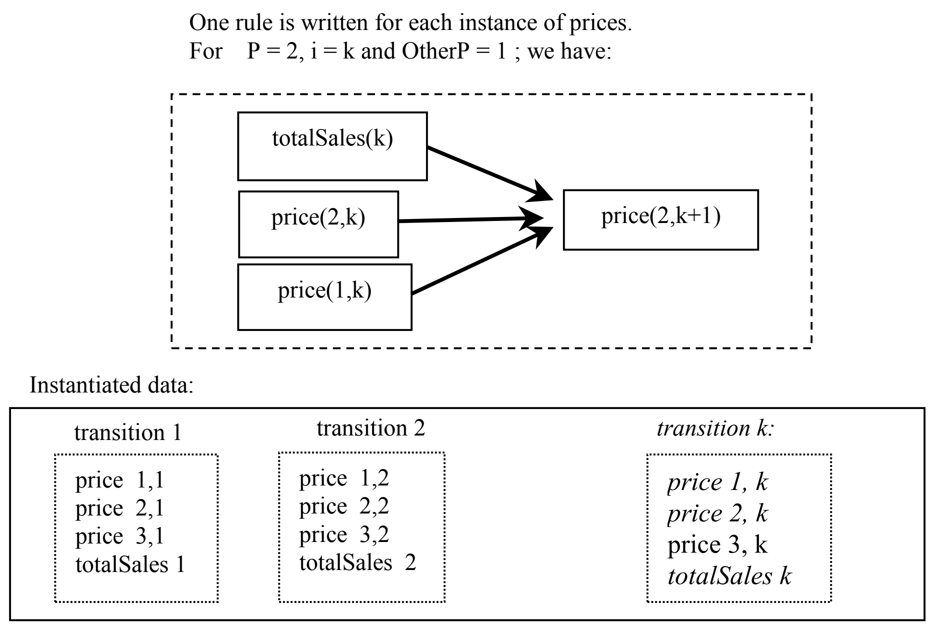


Fig. 6. "Unwrapping" dependencies and instantiate data at iteration i, i = 1, 2, …k

If the name of price is used to make explicit the day, the rule per *day-i* and *producer-j* will have the following form. It is important to observe that *only one instance of newprice in the consequent is associated with only one transition rule and vice verse*:

```
for all (consumer)
(
 price-i(producer-j, myPrice)              and
 totalSales-i(totalSales)                  and
 sales-i(producer-j, mySales)              and
 choiceAnotherProducer(anotherProducer)    and
 price-i(anotherProducer, otherPrice)      and
 calculateNewPrice(mySales,totalSales,
 otherPrice, myPrice,newPrice)
implies
  price-(i+1)(producer-j, newPrice)
```

This process of splitting needs a semantic manipulation of rules allowing explicit reference to instances of data. It is necessary to instantiate data already set up in the database or whose introduction during the simulation can be foreseen. Parameters of the model are examples of data given at the beginning of the simulation. Choices such as selecting another agent in order to compare certain performance parameters and take decisions also can be predicted. This was the case in our application.

Compilation of rules can be implemented in SDML using a meta agent, or *meta module* for our purposes. SDML's meta module can act only at the beginning of the simulation, so its usefulness might be too limited where choices appear during the simulation. In such case a meta module allowing rules to be written at any time during the simulation would make the meta module more powerful and would help even more. Such a meta module might be used, for instance, in an iteration for writing transition rules for the next iteration. Also it might be used to drive the search conveniently, for example, to choose rules to guide the search according to certain criteria, e.g. as the set of support is selected in OTTER. It would also help when the shape of the sought or proved theorem can be adapted during the simulation.

5.2 What the Technique Enables

In this example, the described technique was used to prove that the size of the interval of prices (that is*: biggest price - smaller price,* each day) decreases over time during the first six STIs over a range of one parameterisation and eight choices for the agents at each STI. An exponential decrease of this interval was demonstrated in all the simulation paths. A total of 32768 simulation trajectories were tested. It was not possible to simulate beyond this number of days because of the limitations imposed by computer memory. The complete search process took only 24 hours.

Though the tendency we have shown is simple and quantitative, it is obvious that the technique is applicable in more interesting cases of emergent tendencies, even if they have a qualitative nature.

This technique is useful not only because of the speeding up of the simulation but also for its appropriateness when capturing and proving tendencies under the specified constraints. In the example, the meta-module was used to write the rule with the hypothesis (theorem) to be tested on prover-module at the beginning of the simulation. If the meta-module were able to write rules on prover-module while the simulation is going on, the theorem we wanted to prove could be *adapted* according to the results of the simulation via *relaxing constraints*. For example, the technique could be implemented in a way that we only give the program hints related to the sort of proof we are interested in. Then the meta-module would "*elaborate"*, via adapting over time in a context dependent manner, a set of hypothesis or theorems. That is, a *theorem into a family of alternative theorems will be searched* (those hints will specify such a family) starting from the most constrained one.

6 Other Approaches

Among the more used techniques to win efficiency in simulation we have event-driven simulation using a future event list (useful e.g. in queue simulations) and partition of the space of rules (popular in declarative simulation). SDML uses the second approach. In SDML, the criterion for firing rules is well understood, and procedures like weighting and subsumption used in theorem provers such as OTTER usually are not needed. Additionally, redundant data for some purpose could be avoided in MAS like SDML with appropriate compilation techniques.

The advantages given for the weighting procedure in OTTER are yielded in MAS systems like SDML by procedures such as *partitioning*, where chaining of the rules allows firing the rules in an efficient order according to their dependencies.

Among other approaches for the practical proof of MAS properties, the more pertinent might be the case conducted by people working in DESIRE [5]. They propose the hierarchical verification of MAS properties, and succeeded in doing this for a system.

However, their aim is the verification of a computational program – it is proved that the program behaves in the intended way. It does not include the more difficult task, which we try to address, of establishing general facts about the dynamics of a system when run or comparing them to the behaviour observed in other systems [1].

7 Conclusions and Future Work

We have described and demonstrated a procedure for a constrained exploration of an envelope of trajectories as a complement to traditional methods dealing with post-hoc analysis of the dynamics of simulations. We have suggested a forward chaining generation of trajectories in a semantically constrained way. Constrictions will be context-dependent (over the semantic of the trajectory itself) and will be driven via the introduction of a controller. Once a tendency is identified the idea is to prove its necessity relative to the logic of the simulation language, a range of parameterisations and agents' choices.

A platform to implement this methodology has been proposed. It consists of a modular structure according to strategic parts of a simulation: a first module, *model*, sets up the *static structure* of the simulation; then a second module, *prover*, generates the *dynamics* of the simulation; and finally a *meta-module* is responsible for *controlling* the dynamics of the simulation. The second characteristic of this platform is a *partitioning* of the space of rules and *splitting of transition rules* by STI, parameters and choices.

Allowing a complete model exploration of simulation dynamics for the range of parameters and choices but encoding in a lower information level than that given in MAS, means that a different trade-off and range of facilities is available compared to a MAS and in this sense is complementary. It is at a lower level than a MAS as the hierarchy of agents disappears – the model consists in a single database - rulebase pair. At the new level of programming the modeler will be less informed about individual trajectories but he will get overall information over a complete subset of trajectories corresponding to a range of parameters and choices. Even more, *information about a fragment of the whole theory embedded in the simulation will be collected. This will permit to draw more general conclusions about tendencies as they can be proved for a fragment of the theory.* Moreover, all this is of great interest for the social simulation community as on one hand it helps the modeler to improve his knowledge about and to prove emergent tendencies in a simulation, and on the other hand it can be useful for validation, alignment and simplification of models.

Furthermore, as the exposed technique is model based, it is also useful for scenario exploration. In fact, it can be seen as falling in between inspecting single runs in model exploration of trajectories and syntactic theorem proving. Even lower levels of programming more powerful for proving tendencies but informing less about individual trajectories can be proposed and are a subject for further research.

Acknowledgements

SDML has been developed in VisualWorks 2.5.2, the Smalltalk-80 environment produced by ObjectShare. Free distribution of SDML for use in academic research is made possible by the sponsorship of ObjectShare (UK) Ltd. The research reported here was funded by CONICIT (the Venezuelan Governmental Organisation for promoting Science), by the University of Los Andes, and by the Manchester Metropolitan University Business Faculty.

References

1. Axtell, R., R. Axelrod, J. M. Epstein, and M. D. Cohen, "Aligning Simulation Models: A Case of Study and Results", Computational Mathematical Organization Theory, 1(2) (1996), pp. 123-141.
2. Chiang, C.L., and R. C.T. Lee, Symbolic Logic and Mechanical Theorem Proving, Academic Press, London, UK, 1973
3. Domingo C., G. Tonella and O. Terán, "Generating Scenarios by Structural Simulation", in AI, Simulation and Planning High Autonomy Systems , The Univ. of Arizona (1996) , pp 331-336.
4. Edmonds, B. , "Modelling Bounded Rationality In Agent-Based Simulations using the Evolution of Mental Models". In Brenner, T. (Ed.), Computational Techniques for Modelling Learning in Economics, Kluwer (1999), 305-332.
5. Engelfriet, J., C. Jonker and J. Treur, "Compositional Verification of Multi-Agent Systems in Temporal Multi-Epistemic Logic", AI Group, Vrije Universiteit Amsterdam, The Netherlands, (1998).
6. Konolige, K., "Autoepistemic Logic", in Handbook of Logic in Artificial Intelligence and Logic Programming (4), Oxford Science Publications, 1995, pp. 217-295.
7. McCune, W. (1995), OTTER 3.0 Reference Manual Guide, Argonne National Laboratory, Argonne, IL, 1995.
8. Moss, S., H. Gaylard, S. Wallis, B. Edmonds, "SDML: A Multi-Agent Language for Organizational Modelling", Computational Mathematical Organization Theory, 4(1) (1998), 43-69.

9. Moss, S., B. Edmonds, S. Wallis, "Validation and Verification of Computational Models with Multiple Cognitive Agents", Centre for Policy Modelling Report (1997), CPM-97-25, `http://www.cpm.mmu.ac.uk/cpmrep25.html`
10. Wos, L., Automated Reasoning: 33 Basis Research Problems, Prentice Hall, New Jersey, USA, 1988.
11. Wos, L., Robinson, G. A., and Carson, D. F., "Efficiency and Completeness of the Set of Support Strategy in Theorem Proving", J. ACM 14, N° 4 (1965), 698-709.
12. Zeigler, B., Theory of Modelling and Simulation, Robert E. Krieger Publishing Company, Malabar, Fl, USA, 1976.

Agent-Based Social Simulation with Coalitions in Social Reasoning

Nuno David[1,*], Jaime Simão Sichman[2,**], and Helder Coelho[3]

[1] Department of Information Science and Technology, ISCTE/DCTI, Lisbon, Portugal
Nuno.David@iscte.pt
http://www.iscte.pt/~nmcd

[2] Intelligent Techniques Laboratory, University of São Paulo, Brazil
jaime@pcs.usp.br
http://www.pcs.usp.br/~jaime

[3] Department of Informatics, University of Lisbon, Portugal
hcoelho@di.fc.ul.pt
http://www.di.fc.ul.pt/~hcoelho

Abstract. There is a growing belief that the agents' cognitive structures play a central role on the enhancement of predicative capacities of decision-making strategies. This paper analyses and simulates the construction of cognitive social structures in the process of decision making with multiple actors. In this process it is argued that the agent's rational choices may be assessed by its motivations, according to different patterns of social interactions. We first construct an abstract model of social dependence between agents, and define a set of social structures that are easily identifiable according to potential interactions. We then carry out a set of experiments at micro-social levels of analysis, where the agents' cognitive structures are explicitly represented. These experiments indicate that different social dependence structures imply distinct structural patterns of negotiation proposals, which appear to have diverse patterns of complexity in the search space. It is subsequently shown that this observation emerges as an issue of ambiguity in the regulation of different decision-making criteria, relative to motivation-oriented and utility-oriented choices. In the scope of this ambiguity, we finally make some conjectures relative to further analytical and empirical analysis around the relation between patterns of complexity of social structures and decision-making.

[*] Partially supported by FCT/PRAXIS XXI, Portugal, ref. BM/6876/95 and BD/21595/99.
[**] Partially supported by CNPq, Brazil, grant number 301041/95-4.

S. Moss and P. Davidsson (Eds.): MABS 2000, LNAI 1979, pp. 244-265, 2000.

1 Introduction

The problems encountered in the implementation of autonomous agents that decide and adopt goals on behalf of other agents, have determined a growing need to implement different degrees of social reasoning abilities in the individual agent's machinery [26]. The need for an increasing autonomy in Multi-Agent Systems (MAS) shares some of the difficulties encountered in explanatory models of purposive action [18,20] in the social sciences. These models rest on the assumption that actors in a dynamic social world are purposive, and act in ways that produce intended and/or beneficial results. By advancing the postulate that individual action is goal directed the prevalent question runs around the way people, given their values, beliefs and high-level normative organization behaviors, make choices. The same question naturally arises when designing artificial autonomous agents, and the discipline of Multi-Agent Based Simulation (MABS) naturally emerges as an adequate platform for the study of social reasoning and decision-making strategies in natural or artificial societies. Agents in artificial social systems do not always have control over the other agents' decisions, including the goals they should pursue and the actions they should execute. Such conditions are either constrained by the inherent distribution of goals and knowledge in the system (most problem solving systems using a MAS approach, e.g., [13]) or deliberately defined by the system designer to investigate cognitive aspects of the individual agent and/or emergent properties of the system as a whole (most systems in MABS [14]). In either way, rational autonomous agents need social reasoning abilities to choose goals and partners with adequate capabilities, and to generate proposals to convince the others to collaborate favorably to their collective or individual goals.

The problem of rational choice among a set of feasible alternatives is frequently associated with the question of choice between different decision-making strategies. Some authors advocate a context-bounded notion of rationality, such that different contexts call for different decision-making strategies [7]. For instance, utility theory based on the classic economic principle of rationality does not always conform to human choice behavior [18] and significant evidence in the MABS field seems to show that the ordering of alternatives to maximize the difference between benefits and costs does not provide an increase on the number of coalitions in a multi-strategy world [7]. Even if information is obtained easily and the perfect rationality assumption is relaxed, the individual must often consider alternatives sequentially and decide about them as they are presented. Limited information-processing capacity causes agents to rely on a number of heuristic principles that reduce the complexity of even simple problems, meaning the assumption of utility maximization is discarded for the weaker assumption of procedural rationality [27]. Moreover, there is a growing belief that agents' cognitive and motivational structures play a central role in the enhancement of predicative capacity of decision-making strategies. For instance, this seems to be the main motivation behind the design of the Belief-Desire-Intention (BDI) (e.g.[21]) and the Belief-Values-Goals (BVG) [2] architectures.

There are other attempts to introduce cognitive individual ingredients in the process of decision making, in which, unlike BDI architectures, the social structures between agents are explicitly represented, lending it easily to social simulation based-analysis. In the Theory of Social Power and Dependence Networks [4,26] agents have different capabilities that are complementary to achieve a set of goals. The individual agent behavior is determined by its motivations, according to patterns of social interactions that may occur with other agents, like, for example, social exchange or cooperation. The type of social interactions is determined by the agent's situation in his structure of dependence relations. The notion of rationality is thus based on relational notions of dependence, allowing the definition of different taxonomies of dependence situations between agents. In the present work we will call this type of rationality *motivation-oriented* rationality, and will analyze and simulate the construction of social power and dependence structures [4,26,6,24,8] in the scope of high-level collaboration with generation of proposals for making coalitions with multiple actors. The objective is to analyze the properties of associating motivation-oriented and utility-oriented decision-making criteria in artificial institutions by using multi-agent modeling and multi-agent based simulation.

There are several reasons to account for high-level collaboration models in multi-agent based simulation (MABS) and more generally in multi-agent systems (MAS)[1].

Firstly, agent social interactions frequently occur through high-level communication languages, and consequently are conducted on levels of abstraction within or above Newell's Knowledge Level [19,16]. On a practical level, the system designer usually prescribes the agent's goals. However, the unpredictable nature of the other agents' motivations, and high-level normative organizational behaviors, raises higher the dynamics of the other agents' goals to the eye of the agent. Agents may not only need to exchange tasks or specific actions, but may need to measure, exchange and adopt each other's goals in substantive terms. Secondly, the complexity of social reasoning in terms of goal adoption and goal delegation structures has been shown to be a NP complete problem [10]. Such complexity calls for active experimentation on both micro-social and macro-social levels, in order to assess patterns of interdependencies that may enhance the search for adequate partners and the collaboration process among cognitive agents.

In the scope of this article, we therefore adopt a two step methodological analysis, the first one based on multi-agent modeling and the second on controlled experimentation.

In the first step, we analyze cognitive representations of social dependence structures in the context of relations from a single agent to a non-empty set of agents (1:n). Different power and dependence structures are systematized, conceding different effects in one agent's ability to find and influence others to collaborate. The agents' decision mechanisms use both utility-oriented and motivation-oriented criteria to choose adequate partners and proposals to form coalitions.

[1] An extensive review concerning possible vectors for cross-fertilization among Multi-Agent Systems and Agent-Based Social Simulation may be found in the introductory chapter of the last MABS workshop [14].

In the second step, we use agent-based simulation to test our rationality approach. Here, we advocate that the complexity of social power and dependence patterns may be assessed with the simulation of dependence structures in artificial societies. These simulations may range from highly controlled experiments with emphasis on the individual agent representations of social structures (with an explicit relation to the cognitive agent's machinery) to highly stochastic experiments with a descriptive analysis of the artificial system as a whole (where the relation to the cognitive agent's machinery is more difficult to assess). One objective in our experiments is to emphasize the simulation of cognitive representations of dependence structures at the micro-social level, as being complementary to the simulation and assessment of patterns of dependence at the macro-social level, the last one being usually analyzed in statistical terms.

Perhaps with the exception of Conte and Pedone [7], where the authors try to assess some cognitive ingredients of individual rationality on micro-social and macro-social levels of analysis, one may notice that the literature of MABS [14] has prevalently simulated social phenomena from a macro-social perspective of analysis. This report shares some foundational aspects with [7], namely, that an experimental manipulation of cognitive internal variables is necessary to increase the predicative capacity of decision making and social scientific theories. However, we will restrict our experiments to the micro-social perspective of analysis, and present some further conjectures for future vectors of research that may require us to use a macro-social level of analysis.

We start in section two by presenting a cognitive model of social reasoning that generates different dependence structures and proposals of coalitions with multiple agents. This model is based on a social reasoning mechanism [26,8] and in this paper especially stresses its emphasis on the paradoxical usage of both utility-oriented and motivated-oriented decision-making criteria for selection of partners and generation of proposals. In section three we proceed with the simulation of these representations and present our preliminary results.

The results suggest that distinct dependence situations [26] span different patterns of proposal structures for coalition formation, which seem to have different patterns of complexity in the search space. We further show that such patterns introduce ambiguity in the orderliness of different criteria, related to individual utility-oriented and motivation-oriented decision making. While the agent deliberation dynamics in MAS and MABS calls for combined measures of motivation-oriented and utility-oriented rationality, we suggest that additional analytical work at micro-social levels of analysis and empirical work at macro-social levels of analysis is required, in order to understand and change dynamically the agent's rational abilities according to relations between dependence patterns and the corresponding complexity in the search space.

2 Goal Hierarchies and Adoption

Agents might depend on others (or prefer the others) to achieve some of their goals, which ultimately leads them to negotiate and exchange partially delegated goals. An agent's endogenous goal (e.g. a goal assigned by the system designer) will often need to explore social objects in the exterior world. Strictly speaking, by endogenous we mean a goal that is stored at the Knowledge Level in Newell's sense. Accordingly, new goals (and beliefs) may be acquired in the Knowledge Level owing precisely to the social world. Goals may in fact be adopted instrumentally in order to obtain some advantage in return [4]. If this is the case, the adopted goal may be seen as a "means-to-ends" link to a higher order goal in a tree hierarchy of goals.

We may consider the multiplicity of potential pairs [adopted goal/partner] to be an or-hierarchy sub-tree associated with an agent's endogenous higher-order goal. The question for a rational agent is therefore: which external goals to adopt and to which partners send the corresponding proposals for collaboration? This work does not concentrate on the decision problem related to choice of active endogenous goals, but on the choice of external goals pertaining to such or-hierarchy sub-trees.

In previous work we have proposed a decision model built upon the social reasoning mechanism [26,8], which is based on the Theory of Dependence and Social Power [4]. Shortly, if an agent depends on a third-party agent in order to achieve his goal, the third-party's goals may become candidates for adoption, meaning the adoption is strictly instrumental. The choice of a goal among a set of goal candidates for adoption is based on both quantitative and qualitative measures of dependence relations between the agents.

2.1 Dependence Relations

We consider that the agent $ag_o \in \boldsymbol{Ag}$ is a generic agent in a finite set of agents, designated *subject* agent, who uses his social reasoning mechanism in order to better propose/accept coalition proposals to/from other agents. Agents model the other agents' goals, plans and controlled actions through a data structure that we call *external description*. The external description comprises a finite set of *entries*, each one holding a set of goals, plans and controlled actions for each known agent in the agency. With such a structure a subject agent is able to calculate a set of dependence relations between any specific agent, which here we will call object agent, and his peers[2].

An object agent ag_o is dependent on a third-party agent ag_t, in regard to a specific goal g, according a specific set of plans $\boldsymbol{P}$, iff the object agent needs to execute an action controlled by the third-party agent and not controlled by the object agent - $d\_on(ag_o,ag_t,g,\boldsymbol{P})$. One may have several types of dependence relations among two agents: *unilateral*, *bilateral*, *mutual* and *reciprocal* dependencies. A ***M**utual Dependence* (MD) between the object and the third-party agents represents a bilateral dependence concerning the same goal. A ***R**eciprocal **D**ependence* (RD) defines a bilateral dependence in regard to two different goals. Another concept in the model is the

2 For simplicity and clarity we assume here that the subject and object are the same, i.e., the subject agent reasons about his own properties. We also assume that agents have complete and correct beliefs about each other. These assumptions are not restricted in the social reasoning model, as it may be seen in [25]. Furthermore, we assume that agents are sincere, meaning they do not communicate to others information in which they do not believe.

notion of *dependence situation* (dep-sit), which tries to capture an agent's susceptibility to adopt another agent's goal. Dependence situations relate two agents and a goal, and may be locally or mutually believed, depending on their *source*, i.e. the set of plans that is used to infer them. This is actually a somewhat intuitive notion. For example, let us imagine we are pondering to create a new business company and we are looking for interested partners: it is rather insightful to examine to what extent may we use exclusively our plans to collaborate, meaning the dependence situation is locally believed, or question ourselves if they share an identical opinion, meaning the dependence is mutually believed[3].

In this paper, we will use $\boldsymbol{P}_{ag_o}(ag_o)$ when referring to the object's agent set of plans, and $\boldsymbol{P}_{ag_o}(ag_t)$ when referring to the plans the object agent believes the third-party has. In the latest case we will often abbreviate $\boldsymbol{P}_{ag_o}(ag_t)$ simply to *the third-party agent set of plans*. In addition, we will omit the explicit reference to the object agent in the formulae and will often use $\boldsymbol{P}(ag_t)$ instead of $\boldsymbol{P}_{ag_o}(ag_t)$.

Two elementary relations of dependence called *Inverse Dependence Relations* (IDR) are particularly useful in our work. Each IDR represents a certain amount of power owned by an object agent over a specific third-party agent and goal. Such power may be inferred according to the object agent's set of plans or according to the third-party agent set of plans. We call a third-party agent dependence on the object agent, inferred according to some goal and the plans the object agent thinks the third-party has, a ***R**emote **B**elieved **I**nverse **D**ependence*:

$RBID(ag_o,ag_t,g) \equiv_{def} d\_on(ag_t,ag_o,g,\boldsymbol{P}(ag_t))$.

Conversely, a ***L**ocal **B**elieved **I**nverse **D**ependence* defines a third-party agent dependence on the object agent according to the object agent's set of plans:

$LBID(ag_o,ag_t,g) \equiv_{def} d\_on(ag_t,ag_o,g,\boldsymbol{P}(ag_o))$.

For instance, consider the following airline companies scenario, with an object agent *Af* and his external description shown in figure 1.

Here, goals can be satisfied by flight carriers with desired departure and destination points. Plans represent routes with multiple stops to fulfill multiple market shares.

According to *Af*'s beliefs it is possible to infer that he depends unilaterally on agents *Tp* and *Au*, when considering the goal Paris/Sydney and his own set of plans: *Tp* controls actions Lisbon/Macau and Macau/HongK, and *Au* controls HongK/Sydney. Conversely, agent *Af* may infer a remote believed IDR relative to agent *Tp* and goal Lisbon/Moscow, since *Tp* depends on *Af* for action Paris/Moscow according to *Tp*'s plans. One may also notice that agent *Au* does not originate any IDR according to *Af* beliefs. In fact, *Af* does not have anything to offer to *Au*, either according to *Af*'s plans or *Au*'s plans. The dependence structure in the bottom of the figure identifies agent *Af*'s possible offered goals, plans and actions relative to agent *Tp*. Here, we say that the goal Lisbon/Moscow is an *offered-goal*. The corresponding plan is designated *offered-plan* and the action Paris/Moscow is called an *offered-action*.

[3] More precisely, meaning that we believe that the dependence is mutually believed. We use this notion of mutual belief in the rest of the paper.

Identity: <Af af.somewhere.com 3856> **Goals:** Paris/Sydney(120)
Actions: Paris/Moscow(52); Paris/London(8); Paris/Lisbon(26)
Plans: Paris/Sydney:= Paris/Lisbon, Lisbon/Macau, Macau/HongK, HongK/Sydney.

Identity: <Tp tp.north.com 7352> **Goals:** Lisbon/Moscow(300)
Actions: Lisbon/Paris(26); Lisbon/Macau(156); Macau/HongK(2)
Plans: Lisbon/Moscow:= Lisbon/Paris, Paris/Moscow.

Identity: <Au au.anywhere.com 7366> **Goals:** Sydney/SaoPaulo (45)
Actions: Sydney/BuenosAires(147); HongK/Sydney(100)
Plans: Sydney/SaoPaulo:= Sydney/BuenosAires, BuenosAires/SaoPaulo.

Tp tp.north.com 7352>
|------ Lisbon/Moscow (300) (RBID)
    |------ Lisbon/Moscow:= Lisbon/Paris, Paris/Moscow. (Feasible NLSource)
        |------ Paris/Moscow (52)

Figure 1. An example of dependence relations

In the context of high-level negotiation, any IDR may be seen as a potential proposal to the third-party agent. The object agent has power over the third-party agent desired goal, which is ultimately associated with a set of actions partially controlled by the proponent and some set of plans. We use IDRs to define the set of all possible *offered goals* to the third-party agent.

Formally, the set of offered goals comprises all goals making the third-party agent ag_t dependent on the object agent ag_o, either according to the object agent's set of plans or the third-party agent set of plans, i.e., local or remote believed IDRs:

$$\mathbf{O\text{-}G}(ag_t) \equiv_{def} \{g \in \mathbf{G}(ag_t) \mid LBID(ag_o, ag_t, g) \vee RBID(ag_o, ag_t, g)\}.$$

The corresponding set of possible *offered plans* comprises plans in the object agent's set of plans $\mathbf{P}(ag_o)$ or in the third-party agent set of plans $\mathbf{P}(ag_t)$ for which the third-party agent depends on the object agent[4]:

$$\mathbf{O\text{-}P}(ag_t) \equiv_{def} \{p \in (\mathbf{P}(ag_o) \cup \mathbf{P}(ag_t)) \mid \exists (a \in \mathbf{adep}(ag_o, ag_t), g \in \mathbf{O\text{-}G}(ag_t))(uses(p,a) \wedge goal(p)=g)\}.$$

Finally, the associated set of possible *offered actions* comprises members of the object agent's set of controlled actions $\mathbf{A}(ag_o)$ for which the third-party agent depends according to the set of offered plans and offered goals. Note that an offered action must necessarily be performed by the object agent, although it may be performed according to a plan believed by the object agent and/or believed by the third-party agent:

$$\mathbf{O\text{-}A}(ag_t) \equiv_{def} \{a \in \mathbf{A}(ag_o) \mid \exists (p \in \mathbf{O\text{-}P}(ag_t))\ (uses(p,a) \wedge a \notin \mathbf{A}(ag_t))\}$$

Offered goals are captured by the notion of conjunctive dependencies, namely multi-goal *and*-dependencies, where the third-party depends on the object agent for multiple goals. Conversely, a set of offered plans relative to a same offered goal is captured by the notion of multi-plan *or*-dependencies. In Conte and Castelfranchi [6]

[4] $\mathbf{adep}(ag_o, ag_t)$ is the set of actions controlled by the object agent but not controlled by the third-party.

and David *et al*. [8] it is shown in a substantive sense that conjunctive IDRs augment the power over the third-party, while disjunctive IDRs increases the flexibility for negotiation by augmenting the set of available alternatives.

2.2 Performance, Choice and Rationality

The problem of choice among a set of feasible proposals is inherently connected with expected performance. Generally, if the principle of non-benevolence is assumed, we may find two major trends for measuring the agents' individual performance [5,7,9].

The first one adopts a utility oriented scale, calculated according to the cost of the agents' actions against the worth of the corresponding goals, whatever goals these may be. Such theories specify that when as agent is acting rationally, the agent is engaging in some kind of optimization. The agent's decision functions are fundamentally concerned with the *choice of actions* that maximize utility, often according to the classic principle of economic rationality (e.g.[29]). Choice of goals is not so critical to the individual agent since the agent designer often prescribes (hardwires) the goals in the agent's machinery. Paradoxically, utility-oriented agents may have to drop high value goals in favor of lower value goals if the difference between benefits and costs in the latter case is higher than the former. Also, agents are usually required to have a high level of knowledge and computational ability with which to determine and evaluate a set of available alternatives.

A motivation oriented perspective of individual rationality will most probably value a *substantive* [7,9], hedonistic view, of rationality; that is, individual *performance* measured in terms of the agents' attained goals (e.g. number of goals). Here, similar to Newell's principle of rationality, the real *motive* for being rational is focused on the agents' own goals. In this case, the agent's decision functions are essentially concerned with the choice of adequate partners in order to achieve a set of individual goals. Here, the choice of proper interactions among a set of alternatives is generally qualitative in nature, according to orderings of qualitatively different patterns of dependence between agents.

A number of problems have been identified with classical utility decision theory, like orderability of preferences or computational complexity (e.g.[22]). Nevertheless, these theories seem to be adequate to model a number of social phenomena, such as the problem of emergence of cultural groups [15] or social trade networks [11]. Similarly in MAS with real distributed and open environments (e.g. the Internet), the agents abilities are specified to a great extent in terms of auctions and services (e.g. white and yellow pages, search engines), making utility oriented decision theories adequate to applications such as electronic commerce.

While different utility-oriented models share the fact that agents are purposive, in the sense that they act in ways that tend to produce beneficial results, the heterogeneity of agents and their different goals makes a motivation-oriented notion of performance also desirable. Together with other authors [18,4,7], we advocate that goal directed behavior often results not a from a conscious weighing of the expected future benefits of alternative lines of action, but from a less deliberate response to beliefs internalized through the socializing influences of social structure. For instance, in artificial societies, a crucial operational issue in coalition formation is the problem around the choice of offered goals, selected from a given set of candidate alternative proposals. Another related problem is the issue of delegation and goal adoption, which seems to play a crucial role in human-computer interaction [3]. In a dynamic and heterogeneous world there may be different decision-making strategies to accept coalition proposals, with some agents possibly being more hedonistic and others utilitarian.

These issues ask for complementary types of rationality for the generation of proposals, which in our view must use both utility-driven and motivation-driven strategies. The agents' evaluation of receiving proposals against their goals means that an explicit and social structural link may be established between selection of partners and choice of proposals. In this work we simulate such an approach and utilize the notion of *dependence situations* and *dependence strength*.

The former notion is a motivated-oriented definition of qualitatively different patterns of dependence, calculated according to different configurations of dependence relations between agents.

The later concept has an intended utility and motivation oriented hybrid character, a function expressing the object agent's preferences, with equal probabilities, between actions that may be offered to a same third-party agent. For each possible partner in a coalition, the object agent's *offered action strength* is calculated according to its cost and the substantive contribution to all possible offered goals and plans. This means that for each possible partner there will be a finite set of possible atomic states, each one corresponding to a different action controlled by the object agent. Naturally such a function, which we call *offered action strength*, will often be a domain dependent function. To our ends, we will use the following simplified formula:

$a\text{-}strength_{ag_t}(a) =_{def} (\sum_i N_{plans}(g_i, a).w(ag_t, g_i))/c(ag_o, a))$, where g_i is any offered goal for which the offered action a can contribute, $w(ag_t,g_i)$ is the goal importance according to the third-party agent[5], $N_{plans}(g_i, a)$ is the number of offered plans for goal g_i that use the offered action, and $c(ag_o, a)$ is a positive integer representing the cost of the offered action according to the object agent. Notice that the numerator expresses a hedonistic view of preferences, favoring actions that maximize the contribution to the importance of offered goals.

The notion of dependence strength considers the number of possible offered actions and ponders and integrates their strength:

$$dep\text{-}strength(ag_t) =_{def} \sum_{a \in \mathbf{O\text{-}A}(ag_t)} a\text{-}strength_{ag_t}(a)$$

[5] We assume that the importance of the third-party agent goal is known to the object agent - the computation of the exact importance is in fact not possible in most situations. We however assume that the object agent stores this information in his external description when considering his qualitative knowledge about these goals (e.g. to a certain extent different companies may know each others' order of preferences of strategic goals). Since we do not deal with learning and perception issues in this paper, we do not loose generality in the model and experiences, since their focus is essentially on the properties of social dependence networks and its cognitive representations.

The latter definition identifies the most dependent agents on the proponent according to the relevance of his set of available proposals. The former formula suggests the most valued offered actions, playing an important role during the selection of negotiation proposals.

2.3 Choice of Partners and Proposals

Suppose that some object agent ag_o is pursuing some goal g_e and commits to some plan p_e called respectively the *engaged goal* and *engaged plan*. Let us assume he is dependent on others to achieve that goal and execute that plan. Also, for every action a^d on which the agent depends on others in the plan p_e, there is a non-empty set of possible partners represented in the external description that are able to perform it (i.e. the plan is feasible [23]). Furthermore, possibly different patterns of dependence relations will hold for each possible partner.

If the object agent depends on a possible partner for the engaged goal and plan, he may wish to calculate if the latter also depends on him for some of his goals and plans. However, their set of plans may differ, and the object agent may infer, for instance, a mutual dependence relating him and a possible partner, whereas the latter does not infer the same bilateral dependence according to his plans. In order to capture this possible awareness of the partners, a notion called *dependence situation* was defined [26]. In the rest of this paper we adopt the taxonomy and partial ordered set of dependence situations used in [24]: MBMD > (MBRD,LBMD) > LBRD > UD (meaning for instance that MBMD is *higher* than UD).

The last two letters in the acronyms differentiate Mutual Dependencies (MD) from Reciprocal Dependencies (RD). As for the first two letters, if the dep-sit is Mutually Believed (MB) it indicates that is inferred according to both the object and the third-party set of plans. If the dep-sit is Locally Believed (LB) indicates that the dep-sit is inferred according to the object agent's set of plans. UD stands for Unilateral Dependence, meaning the object agent depends on the third-party but the latter does not depend on the former according to the object agent's set of plans, i.e., there are no LBIDs. There is a minor difference here from [24] in that we do not use the situation named IND (Independence), since we assume the object agent depends on others for the engaged goal.

Consider the function $dep\text{-}sit_{ag_o}(ag_t,g_e)$ that calculates the dependence situation according to the object agent, a third-party agent $ag_t \in \boldsymbol{Ag}$ and the object agent' engaged goal $g_e \in \boldsymbol{G}(ag_o)$. We next describe a collection of partial ordered sets and decision functions with respect to the choice of possible partners and the corresponding proposals.

Choice of Partners - if two agents pertain to a same set of possible partners for the object agent, then $ag' \leq_{parrtner} ag$ iff: (**1**) $ag'=ag$; or (**2**) if ag' dep-sit regarding ag_o and g_e is lower then ag; or (**3**) agents have equal deps-sits and ag' dependence strength is lower than ag; or (**4**) agents have equal dep-sits, equal dependence strengths and the cost of the action a^d according to ag' is higher than ag.

In conclusion, for each action the preferred partner is chosen from the corresponding set of possible partners according to a sequence of priorities, primarily motivation-driven (first and second criteria), but also utility-driven (second and third criteria).

Choice of Offered Goals - Except for unilateral dependencies, the set of *chosen offered goals* for each preferred partner, results primarily from the set of offered goals originating the highest dep-sit.

$$\textbf{C-OG}(ag_t,g_e)\equiv$$

$$\{g_e\} \qquad \text{if } dep\text{-}sit_{ag_o}(ag_t,g_e)=MBMD \text{ or } LBMD$$

$$\{g'\in \textbf{O-G}(ag_t) \mid LBID(ag_o,ag_t,g') \wedge RBID(ag_o,ag_t,g')\} \qquad \text{if } dep\text{-}sit_{ag_o}(ag_t,g_e)=MBRD$$

$$\{g'\in \textbf{O-G}(ag_t) \mid LBID(ag_o,ag_t,g')\} \qquad \text{if } dep\text{-}sit_{ag_o}(ag_t,g_e)=LBRD$$

$$\{g'\in \textbf{O-G}(ag_t) \mid RBID(ag_o,ag_t,g') \qquad \text{if } dep\text{-}sit_{ag_o}(ag_t,g_e)=UD$$

Notice in the case of unilateral dependencies (UD), that all chosen offered goals result necessarily from the set of plans the object agent thinks the preferred partner has, i.e., Remote Believed IDRs. In the case of mutual dependencies the engaged goal and the chosen offered goal are necessarily the same.

Choice of Offered Plans - Similar to the computation of offered goals, the set of *chosen offered plans* is highly dependent on the inferred dep-sit. The best feasible offered plans are the ones believed by both agents. Local believed plans are also preferred to non-local believed plans.

Consider a set of offered plans calculated according to the chosen offered goals. The *chosen offered plans* – $\textbf{C-OP}(ag_t,g_e)$ – are calculated according to the following partial order: $p1 \leq_{plan} p2$: iff (**1**) $p1=p2$ or $p1$ is not feasible and $p2$ is feasible; or (**2**) both are feasible and $p1$ is not mutually believed and $p2$ is mutually believed; or (**3**) both plans are feasible and $p1$ is not locally believed and $p2$ is locally believed.

Choice of Offered Actions - Consider a set of offered actions calculated according to the chosen offered plans. The set of *chosen offered actions* – $\textbf{C-OA}(ag_t,g_e)$ – are the ones calculated according to the chosen offered plans and sharing the highest dependence strength.

In summary, the preferred offered action is chosen from the object agent's set of controlled actions associated with (1) offered goals originating the highest dep-sit; (2) feasible and convenient source set of plans; (3) the maximum observed action strength. Formally, the final proposal for each preferred partner ag_t, relative to the object agent's engaged goal g_e is therefore:

$decide_{prop}(ag_t,g_e)=_{def}(a, \textbf{P}(a), \textbf{G}(a))$, with,

$a=random(\textbf{C-OA}(ag_t,g_e))$,

$\textbf{P}(a)=\{p \in \textbf{C-OP}(ag_t,g_e) \mid uses(p,a)\}$

$\textbf{G}(a)=\{g \in \textbf{C-OG}(ag_t,g_e) \mid \exists(p \in \textbf{P}(a))\ (goal(p)=g)\}$.

Strong offered actions are likely to cause positive social interference with several offered plans and goals, increasing the quality of a proposal and the preferred partner's susceptibility to accept the coalition.

3 Experimentation

Social simulation was the way to evaluate our ideas and find predominant patterns of dependence that may be better accommodated in the model. We have implemented short experiments for e-contracts with software packages for reuse [8] and more extensive experiments for strategic reasoning with airline transportation carriers. The latter example, which we will present here, is a typical domain where companies may establish coalitions in order to increase the number of carriers and destinations, for instance, when building packages of lower price flights between multiple cities that one individual company can not provide.

The experiments proceed in small steps and are highly controlled, with an almost absence of random variables. The classical social simulation approach, inspecting over emergent phenomena on a macro-social level (usually described in statistical terms) is not our aim here. This would in fact be a difficult task as all objects (goals, plans and actions) have a clear semantics, and are not randomly generated. Furthermore, and to a certain extent, the model itself shapes the relations that agents are allowed to establish. Following [24], we therefore adopt a lower level analysis and try to proceed slowly for an incremental understanding of social structures created by deliberative agents.

3.1 A First Simple Example

Companies must have a number of common goals and cross dependent carriers so as to make an effective strategic agreement. Goals are available or desired carriers. Each company ascribes a certain importance to their goals. Plans represent routes with multiple stops to fulfill multiple market shares. There may be several plans for a same carrier and each company has its own set of preferred plans. Initially, suppose that there are two agents known to agent *Af*: the agents *Tp* and *Au.* In figure 2 we show the external description of agent *Af*.

I'm agent Af, running at af.somewhere.com, with pid 3856.

Identity: <Af af.somewhere.com 3856>
Goals: Paris/Sydney(120); Paris/Dublin(116); Rome/Boston(40); Rome/Marseille(33)
Actions: Paris/Moscow(52); Paris/London(8); Paris/Lisbon(26); Paris/Argel(22); Paris/Marseille(6); Argel/Dackar(22); Paris/NewY(102); Paris/Toulouse(5); Toulouse/Marseille(6)
Plans: Paris/Dublin:= Paris/London, London/Dublin.
Paris/Sydney:= Paris/London, London/HongK, HongK/Sydney.
Paris/Sydney:= Paris/Lisbon, Lisbon/Macau, Macau/HongK, HongK/Sydney.
Rome/Boston:= Rome/Paris, Paris/NewY, NewY/Boston.
Rome/Marseille:= Rome/Paris, Paris/Toulouse, Toulouse/Marseille.

I have received the following messages of introduction:

Identity: <Tp tp.north.com 7352> **Goals:** Lisbon/Moscow(300)
Actions: Lisbon/Paris(26); Lisbon/Macau(156); Macau/HongK(2)
Plans: Lisbon/Moscow:= Lisbon/Paris, Paris/Moscow.

Identity: <Au au.anywhere.com 7366> **Goals:** Sydney/SaoPaulo (45)
Actions: Sydney/BuenosAires(147); Sydney/Pretoria(156); HongK/Sydney(100)
Plans: Sydney/SaoPaulo:= Sydney/BuenosAires, BuenosAires/SaoPaulo.

Figure 2. External description of agent *Af*

In figure 3, agent *Af* builds his dependence network. We will present hereafter two different kinds of networks. The first type, called *dependence network*, is constructed when the agent reasons about his goals and expresses in a same structure all the object agent's needed actions, considering all of his goals and plans. The second one, called *proposal network*, is constructed when reasoning about partners and presents all possible partners and possible proposals for each needed action in some engaged plan.

In the dependence network, it can be observed that the agent has two goals (Paris/Sydney and Rome/Marseille). However, agent *Af* has only one achievable goal – Paris/Sydney. He has two plans for that goal, but only one of them is feasible. Even though the plan passing by Macau and Hong Kong might not be the most advantageous plan, there are no other feasible plans for his goal and he will try to form a coalition. He is dependent on agent *Tp* for two needed actions – Lisbon/Macau and Macau/HongK – and on agent *Au* for needed action HongK/Sydney.

The information that agent *Af* captures from the proposal network is related to the goals, plans and actions that he can offer to his possible partners. For each needed action and possible partner in the engaged plan we may find information concerning the highest dep-sit (d-sit), the dependence strength on the object agent (d-strength), and the cost ascribed by the possible partner to the needed action (d-a-cost).

```
========== Reasoning about goals ...
My dependence network is:
 <Af>
-- Paris/Sydney (120) (achievable)
  |---- Paris/Sydney:= Paris/London, London/HongK, HongK/Sydney. (NFeasible)
  |    |---------- London/HongK (NA)
  |    |           |********** UNKNOWN
  |    |           | HongK/Sydney (EC:100.0)
  |    |           |********** <Au 7366> (100.0)
 ( . . . )
  |    | Paris/Sydney:= Paris/Lisbon, Lisbon/Macau, Macau/HongK, HongK/Sydney. (EC:284)
  |    |---------- Lisbon/Macau (EC:156.0)
  |                |**********  <Tp 7352> (156.0)
  |                | Macau/HongK (EC:2.0)
  |                |**********  <Tp 7352> (2.0)
  |                | HongK/Sydney (EC:100.0)
  |                |********** <Au 7366> (100.0)
 ( . . . )
  | Rome/Marseille (33) (not achievable)
  |---- Rome/Marseille:= Rome/Paris, Paris/Toulouse, Toulouse/Marseille. (NFeasible)
       |---------- Rome/Paris (NA)
                   |---------- UNKNOWN
                               |----------
The engaged goal is: Paris/Sydney (120)
The engaged plan is:
  Paris/Sydney:= Paris/Lisbon, Lisbon/Macau, Macau/HongK, HongK/Sydney. (EC:284)
========== Reasoning about partners ...
My needed actions are: <Lisbon/Macau>, <Macau/HongK>, <HongK/Sydney>
My possible partners, offered goals, plans and actions for each needed action are:
 Lisbon/Macau and Macau/HongK
```

```
|-- <Tp tp.north.com 7352> / d-sit: UD / d-strength: 5.4 / d-a-cost: 156.0 and 2.0
  |------- Lisbon/Moscow (300) (RBID)
          |------- Lisbon/Moscow:= Lisbon/Paris, Paris/Moscow. (Feasible NLSource)
                  |------ Paris/Moscow (52)
                          |------
 HongK/Sydney
|-- <Au au.anywhere.com 7366> / d-sit: UD / d-strength: 0.0 / d-a-cost: 100.0
  |-------- no offered goals
```

Figure 3. Dependence and proposal network of agent *Af*

In this scenario, all inferred dep-sits are Unilateral Dependencies (d-sit=UD). However, agent *Af* has in fact something to propose to agent *Tp*. The power of *Af* over *Tp* is not insignificant, according to the plans *Af* thinks *Tp* has: agent *Af* may be able to use in an instrumental way the Remote Believed IDR with his proposal involving the offered goal Lisbon/Moscow and the offered action Paris/Moscow. On the contrary, agent *Au* dependence strength on *Af* is zero. Agent *Af* will not propose anything to agent *Au* as shown in figure 4.

In figure 4, agent *Af* receives *Tp*'s acceptance of proposal. On the other hand, agent *Au* will reject *Af*'s proposal, which is justified by the non-benevolence principle. In reality nothing was proposed to *Au*. The plan is no longer feasible since there are no more possible partners available and all the agent's goals become non-achievable.

The coalition was not formed. This example demonstrates on a practical level is that it seems intuitive to specialize Sichman's dependence situations on both qualitative and quantitative levels. A same dependence situation may be associated with different influencing power conditions. Zero dependence strength implied scarcity of substantive arguments to offer to agent *Au*. Yet, unilateral dependencies inferred according to the proponent's plans, with non-zero dependence strength, may be reciprocated - it is also a priority to search for relations of power on the others agents' beliefs. This was the case of agent *Tp*. Accordingly, it seems clear that social exchange [4] may be triggered by unilateral dependencies coupled with remote believed IDRs (e.g. *Af* with *Tp*), at least if not adopting pure cognitive-psychological examples like in [24].

```
========== Deciding about partners ... (Partner choices criteria: d-sit > d-strength > action_cost )
The selected partner(s) and proposal(s) are :
| Needed actions: Lisbon/Macau, Macau/HongK                | Needed action: HongK/Sydney
| Partner: <Tp tp.north.com>                               | Partner: <Au au.anywhere.com>
| Offered goal/action:<Lisbon/Moscow>/<Paris/Moscow>       | Offered goal/action: NONE/NONE
========== Sending and receiving messages ...
Sending proposals of coalition to <Tp tp.north.com 7352> ... <Au au.anywhere.com 7366>
The messages received are: (Acceptance <Tp 7352>), Refusal <Au 7366>)
My new list of possible partners, offered goals, plans and actions is:
 HongK/Sydney
|------ no possible partners (empty list)
Informing agent <Tp tp.north.com 7352> that the proposal of coalition was canceled ...
========== Reasoning about goals ...
The engaged goal is no longer achievable.
```

Figure 4. Selection of partners and proposals

3.2 Second Example

Let us suppose that after the previous events, four agents arrive at the agency: *Ba*, *Tw*, *Ai1* and *Ai2*. Additionally, to save space, suppose that agents *Tp* and *Au* had left the agency. In figure 5 we show *Ai1*, *Ai2*, *Ba* and *Tw* external description entries.

In this scenario, the autonomy of agent *Af* increases significantly, with all his goals becoming achievable - figure 6. Still, his most important goal will be the same - Paris/Sydney - even though he will choose another plan due to feasibility conditions. There are two needed actions in this plan, London/HongK and HonK/Sydney. In the dependence network represented in the figure, all four agents are able to execute *Af*'s needed action HongK/Sydney.

I'm agent Af, at af.somewhere.com, with pid 3856, I have received the following messages of introduction:

Identity: <Ai1 ai1.anywhere.com 3855>
Goals: Rome/Boston(55); Rome/Marseille(34); Rome/London(20)
Actions: HongK/Sydney(113); Rome/Lisbon(23); Rome/Paris(12)
Plans: Rome/Boston:= Rome/Lisbon, Lisbon/NewY, NewY/Boston.
Rome/Boston:= Rome/Paris, Paris/NewY, NewY/Boston.
Rome/Boston:= Rome/Paris, Paris/London, London/NewY, NewY/Boston.
Rome/Marseille:= Rome/Paris, Paris/Marseille. Rome/London:= Rome/Paris, Paris/London.

Identity: <Ai2 ai2.somewhere.com 3860> **Goals:** Rome/Boston(55)
Actions: HongK/Sydney(113); Rome/Lisbon(23); Rome/Paris(12)
Plans: Rome/Boston:= Rome/Lisbon, Lisbon/NewY, NewY/Boston.
Rome/Boston:= Rome/Paris, Paris/NewY, NewY/Boston.

Identity: <Ba ba.somewhere.org 3861> **Goals:** London/Maputo(100); London/Argel(90)
Actions: London/Paris(8); London/HongK(158); HongK/Sydney(113); London/Dublin(4)
Plans: London/Argel:= London/Paris, Paris/Argel.

Identity: <Tw tw.air.org 3865>
Goals: NewY/Argel(67); NewY/Camberra(60); NewY/Marseille(64); NewY/Dackar (66)
Actions: HongK/Sydney(103); NewY/London(115); NewY/Paris(120); NewY/Boston(10)
Plans: NewY/Argel:= NewY/London, London/Paris, Paris/Argel.
NewY/Dackar:= NewY/Paris, Paris/Argel, Argel/Dackar.
NewY/Argel:= NewY/Paris, Paris/Argel. NewY/Marseille:= NewY/Paris, Paris/Marseille.

Figure 5. *Ai1*, *Ai2*, *Ba* and *Tw* entries in the external description of agent *Af*

With respect to the action HongK/Sydney shown in figure 7, both agents *Ai1* and *Ai2* share a higher dep-sit - Mutual Believed Reciprocal Dependence (MBRD) - than the one originated by *Tw* and *Ba* - Unilateral Dependence (UD). However, *Ai1*'s dependence strength on *Af* (=17.2) is higher compared to *Ai2* (=0.5), giving to *Af* a significant potential flexibility to negotiate with *Ai1*. For example, *Af* is aware that the action Paris/London may be useful for two of *Ai1*'s current goals (Rome/Boston and Rome/London). Also, notice that agents *Ai1*, *Ai2* and *Ba* share the same and the highest cost (d-a-cost=113.0) for the referred needed action HongK/Sydney. Yet, in figure 8, the strategic choice to execute the needed action will not fall on agent *Tw*, which assigns the lowest cost to the needed action but originates the lowest dep-sit (UD).

As shown in figure 8, agent *Ai1* was thus selected to execute the needed action HongK/Sydney. The possible chosen offered goals are the ones originating the highest dep-sit - Rome/Boston and Rome/Marseille (MBRD). The final choice of proposals to *Ai1* – Rome/Boston as the offered goal and Paris/NewY as the offered action – holds some subtle points: (1) the action Paris/London, although less expensive, belongs to a non-feasible local believed plan (NFeasible, LSource) – there would be no apparent reason for *Af* to send this proposition; (2) even though actions Paris/Toulose and Toulouse/Marseille appertain to feasible plans, they are solely associated with *Af*'s local believed plans (LSource) – there would be no apparent reason for *Ai1* to accept such propositions; and (3) action Paris/Marseille is associated with a non-local believed plan (NLSource) – although it may be possible that the partner would be willing to accept it, there is one other plan believed by both sources that seems to be a better choice for *Af*.

```
========= Reasoning about goals ...
My dependence network is:
 <Af>
-- Paris/Sydney (120) (achievable)
  |------ Paris/Sydney:= Paris/London, London/HongK, HongK/Sydney. (EC:276.5)
  |     |---------- London/HongK (EC:158.0)
  |     |           |**********  <Ba 3861> (158.0)
  |     |           | HongK/Sydney (EC:110.5)
  |     |           |**********  <Ai1 3855> (113.0), <Ai2 3860> (113.0), <Ba 3861> (113.0), <Tw 3865> (103.0)
 ( . . . )
  | Paris/Dublin (116) (achievable)
  |------ Paris/Dublin:= Paris/London, London/Dublin. (EC:12.0)
  |     |---------- London/Dublin (EC:4.0)
  |                |----------   <Ba 3861> (4.0)
  |                         |----------
  | Rome/Boston (40) (achievable)
  |------ Rome/Boston:= Rome/Paris, Paris/NewY, NewY/Boston. (EC:124.0)
  |     |---------- Rome/Paris (EC:12.0)
  |                |**********   <Ai1 3855> (12.0), <Ai2 3860> (12.0)
  |                | NewY/Boston (EC:10.0)
  |                |**********   <Tw 3865> (10.0)
 ( . . . )
The engaged goal is: Paris/Sydney (120)
The engaged plan is: Paris/London, London/HongK, HongK/Sydney. (feasible) (276.5)
```

Figure 6. Dependence network of agent *Af*

```
========= Reasoning about partners ...
My needed actions are: <London/HongK>, <HongK/Sydney>
My possible partners, offered goals, plans and actions for each action are:
London/HongK
|-- <Ba ba.somewhere.com 3861> / d-sit: UD / d-strength: 4.1 / d-a-cost: 158.0
   |-- London/Argel (90) (RBID)
      |--- London/Argel:= London/Paris, Paris/Argel. (Feasible NLSource)
         |--------- Paris/Argel (22)
HongK/Sydney
|- <Ai1 ai1.anywhere.com 3855> / d-sit: MBRD / d-strength: 17.2 / d-a-cost: 113.0
  |---  Rome/Boston (55) (MBRD)
```

```
|   |--  Rome/Boston:= Rome/Paris, Paris/NewY, NewY/Boston. (Feasible BSources)
|   |   |--------- Paris/NewY (102)
|   |   |           |------
|   |   | Rome/Boston:= Rome/Paris, Paris/London, London/NewY, NewY/Boston. (NFeasible
|   |   |--------- Paris/London (8)                                            NLSource)
|   |               |------
|   | Rome/Marseille (34) (MBRD)
|   |--  Rome/Marseille:= Rome/Paris,Paris/Toulouse,Toulouse/Marseille.(Feasible LSource)
|   |   |--------- Paris/Toulouse (5), Toulouse/Marseille (6)
|   |   |           |-------
|   |   | Rome/Marseille:= Rome/Paris, Paris/Marseille. (Feasible NLSource)
|   |   |--------- Paris/Marseille (6)
|   |               |-------
|   | Rome/London (20) (RBID)
|   |--  Rome/London:= Rome/Paris, Paris/London. (Feasible NLSource)
|       |--------- Paris/London (8)
|                   |------
| <Ai2 ai2.somewhere.com 3860> / d-sit: MBRD / d-strength: 0.5 / d-a-cost: 113.0
|---- Rome/Boston (55) (MBRD)
|   |-- Rome/Boston:= Rome/Paris, Paris/NewY, NewY/Boston. (Feasible BSources)
|       |--------- Paris/NewY (102)
|                   |------
| <Ba ba.somewhere.com 3861> / d-sit: UD / d-strength: 4.1 / d-a-cost: 113.0
|---- London/Argel (90) (RBID )
|   |-- London/Argel:= London/Paris, Paris/Argel. (Feasible NLSource)
|       |--------- Paris/Argel (22)
|                   |-----
| <Tw .air.org 3865> / d-sit: UD / d-strength: 95.4 / d-a-cost: 103.0
|---- NewY/Argel (67) (RBID)
    |--  NewY/Argel:= NewY/London, London/Paris, Paris/Argel. (Feasible NLSource)
    |  | NewY/Argel:= NewY/Paris, Paris/Argel. (Feasible NLSource)
    |  |--------- Paris/Argel (22)
    |             |-----
    | NewY/Marseille (64) (RBID)
    |-- NewY/Marseille:= NewY/Paris, Paris/Marseille. (Feasible NLSource)
    |  |--------- Paris/Marseille (6)
    |             |-------
    | NewY/Dackar (66) (RBID)
    |-- NewY/Dackar:= NewY/Paris, Paris/Argel, Argel/Dackar. (Feasible NLSource)
       |--------- Paris/Argel (22), Argel/Dackar (22)
```

Figure 7. Proposal network of agent *Af*

```
========== Deciding about partners ... (Partner choices criteria: d-sit > d-strength > action_cost)
The selected partner(s) and proposal(s) are:
¦ Needed action: HongK/Sydney, Partner: <Ai1 ai1.anywhere.com>
¦ Offered goal/action: <Rome/Boston>/<Paris/NewY>
¦ Needed action: London/HongK, Partner: <Ba af.somewhere.com>
¦ Offered goal/action: <London/Argel>/<Paris/Argel>
```

Figure 8. Selection of partners and generation of proposals

In effect, the action Paris/NewY is associated with a plan believed by both sources (BSources) and there is a mutual interest to form a coalition associated with that plan and goal. This is actually true and in figure 9 the proposal networks of agent *Ai1* shows that there is in fact a Mutual Believed Mutual Dependence (MBMD) relative to the proponent *Af* and to the goal Rome/Boston that is the most valuable for *Ai1*.

Agent *Ai1* accepts the proposal since he needs in fact the proposed offered goal and offered action, as shown in figure 9. This mutual dependence arises since *Af* has the goal Rome/Boston as well. Nevertheless, it differs from *Af*'s dep-sit relative to *Ai1* since they do not share the same set of plans.

```
   ========== Reasoning about messages ...
I have received a proposal of coalition: (PROPOSAL <Af af.somewhere.com 3856>
========== Reasoning about plans ...
My dependence network with reference to the proposed goal <Rome/Boston> is:
 <Ai1>
-- Rome/Boston (55) (achievable)
  |---Rome/Boston:= Rome/Lisbon, Lisbon/NewY, NewY/Boston. (NFeasible)
     |---------- Lisbon/NewY (NA)
     ( . . . )          ( . . . )
     | Rome/Boston:= Rome/Paris, Paris/NewY, NewY/Boston. (EC:124.0)
     |---------- Paris/NewY (EC:102.0)
     |          |********** <Af 3856> (102.0)
     |          |           |----------
     |          | NewY/Boston (EC:10.0)
     |          |**********  <Tw 3865> (10.0)
     |                      |----------
     | Rome/Boston:= Rome/Paris, Paris/London, London/NewY, NewY/Boston. (NFeasible)
     |---------- Paris/London (EC:8.0)
                |********** <Af 3856> (8.0)
                |           |----------
                | London/NewY (NA)
               ( . . . )
The engaged plan is: Rome/Paris, Paris/NewY, NewY/Boston. (feasible) (124.0)
My possible partners, offered goals, plans and actions are (action Paris/NewY):
Paris/NewY
|--
  | <Af af.somewhere.com 3856> / d-sit: MBMD / d-strength: 88.0 / d-a-cost: 102.0
  |---- Rome/Boston (40) (MBMD)
      |----- Rome/Boston:= Rome/Lisbon, Lisbon/NewY, NewY/Boston. (NFeasible LSource)
      |     |-------- Rome/Lisbon (23)
      |     |         |-----
      |     | Rome/Boston:= Rome/Paris, Paris/NewY, NewY/Boston. (Feasible BSources)
      |     |-------- Rome/Paris (12)
     ( . . . )
I will accept the proposal, because I do not have a better partner.
```

Figure 9. Dependence and proposal network of agent *Ai1*

Agent *Ba* accepts *Af*'s proposal as well and the coalition is formed (not shown here). Similar to the first experiment, the proponent *Af* uses his bargaining power over

Ba according to the plans he thinks *Ba* has (using *Ba*'s goal London/Argel and needed action Paris/Argel).

3.3 Some Preliminary Comments - Substance, Utility and Complexity

The model seems to present a coherent behavior, however there are some comments to be pointed out. For instance, a critical issue concerns the use of the decision-making criteria for choice of partners, in which the motivation oriented criteria (dependence situations) deliberately preceded the combined utility/motivation-oriented criteria (dependence strength). A closer look to the second example would show that a change of priorities in the partner selection criteria would elect agent *Tw* instead of *Ai* for the needed action HongK/Sydney. Agent *Tw* originates a lower dependence situation (UD) than agent *Ai1* (MBRD). That is, the experiments give no clue whether high dependence situations with small dependence strengths should be preferred to low dependence situations with high dependence strengths.

A solution to this problem may be possible if a relation between dependence situations and some kind of expected dependence strength can be predicted. For example, let us admit the possibility of a higher trust between partners pursuing a same goal, against the case of pursuing different goals [4,24]. Accordingly, we have considered Mutual Dependencies (one single offered goal) more valuable than Reciprocal Dependencies (one or more offered goals). Nevertheless it is possible to observe, after running multiple experiments, that MDs seem to contribute less than RDs to the overall network dimension and thus overall dependence strength.

A similar point can be noticed with respect to the locality of plans: Mutual Believed Reciprocal Dependencies usually exhibit lower network dimensions than Local Believed Reciprocal Dependencies, since the number of plans in the network seems to be reduced with the mutual believed case, due to the intersection set of remote and local believed IDRs (see section 2.3).

These issues raise some questions in terms of patterns of complexity in the search space. For instance, mutual dependencies often offer a constrained expected space of search for alternatives due to lower network dimensions. It is well known that agents have bounded rationality [27]. We thus may ask: when should trust on a reduced set of alternatives (e.g. mutual dependencies) be preferred to complex, but flexible, networks of possible proposals and bargaining power, eventually with better expected utility benefits (e.g. reciprocal dependencies)? Also, unlike our model of hybrid rationality that confronts both decision approaches in the same level of abstraction, should utility oriented analysis be analyzed on a distinct level of abstraction from the motivation perspective, like for example in [1]? Interestingly, the software engineering oriented work of Jennings and Campos [17] raises higher a set of utility oriented principles to the Social Level of abstraction, above Newell's Knowledge Level.

Complexity can certainly get worse as may be observed at the end of the second experiment. Here, the partner's Mutual Believed Mutual Dependency (*Ai1*) relative to the proponent (*Af*) plays an important role for the coalition settlement, suggesting that it could be interesting for the proponent to analyze the dependence situations relating himself to the other agents. The case would get farther worse if transitive relations were analyzed, leading to analysis of group cohesion.

Regarding these latter points, our experiences are not yet satisfactory. Our conjecture is that further effort is needed to account for a clear relation between complexity and frequently observed dependence patterns. Such effort must be made, on one hand, with further analytical considerations on the complexity of dependence structures (micro-social level) and, on another hand, with empirical analysis of patterns of dependence structures (macro-social level). If such a relation is established, then a dynamic readjustment between utility-oriented and motivation-oriented rationality may be better achieved, by adapting dynamically the agents' rational abilities according to the complexity associated with each dependence situation and available resources. We would also add that this observation clearly establishes methodological evidences around the complementary character of (i) analytical and empirical considerations on the complexity of cognitive social structures at micro-social levels and (ii) empirical analysis of patterns of social dependence structures at macro-social levels.

4 Conclusions

While the technology and normative references for agent interoperability in MAS (e.g. [12]) and MABS (e.g. SWARM [28]) is rapidly being deployed in a wide range of platforms, the predicative tools to deal with complex patterns of social dependencies that emerge within and between agent artificial societies are still inadequate. This work followed a two step methodological approach involving modeling and social simulation based-analysis. Our model for coalition formation was built on the assumption that dependence based choices of partners and proposals are obligatorily integrated issues. The experiments were accomplished at the micro-social level and were able to identify different degrees of influencing power for a same dependence situation, suggesting that social exchange may also be triggered by Unilateral Dependencies.

Further experimental results indicated that different dependence situations span different patterns of proposal structures, which appear to have different degrees of complexity in the search space. This observation emerged as an issue of ambiguity in the previous underlined decision model, concerning the orderliness of different criteria with respect to motivation-oriented and utility-oriented choices. While MAS dynamics calls for combined measures of motivation and utility oriented rationality, we claim that additional analytical analysis at micro-social levels and empirical analysis at macro-social levels is required. Such analysis may open the way to disambiguate and dynamically affect the agents' rational abilities according to relations between patterns of dependence and expected complexity in the search space.

References

1. Alonso E. (1998), *How Agents Negotiate Societies*, Proceedings of the 3rd International Conference on Multi-Agent Systems, ICMAS'98, IEEE press, pages 18-29.
2. Antunes L., Faria J. and Coelho H. (2000), *Improving Choice Mechanisms within the BVG Architecture*, The International Workshop on Agent Theories, Architectures, and Languages (ATAL00), Intelligents Agents VII, Springer Verlag.
3. Castelfranchi C. and Falcone R. (1999), *Towards a Theory of Delegation for Agent-Based Systems*, Robotics and Autonomous Systems, vol. 24, N.3 and 4.
4. Castelfranchi C., Miceli M. and Cesta A. (1992), *Dependence relations among autonomous agents*, Proceedings of MAAMAW92, Elsevier Science, pages 215-227.
5. Coleman J. and Fararo T. (1992), *Rational Choice Theory - Advocacy and Critique*, Sage Publications, Key Issues in Sociological Theory.
6. Conte R. and Castelfranchi C. (1996), *Simulating multi-agent interdependencies. A two-way approach to the micro-macro link*, Microsimulation and the Social Science, Mueller U. and Troitzsch K. editors, Lecture Notes in Economics, Springer Verlag, pages 394-415.
7. Conte R. and Pedone R. (1998), *Finding the best partner: The PART-NET system*, Multi-Agent Systems and Agent-Based Simulation, Proceedings of MABS98, Gilbert N., Sichman J.S. and Conte R. editors, LNAI1534, Springer Verlag, pages 156-168.
8. David N., Sichman J.S. and Coelho H. (1999), *Extending social reasoning to cope with multiple partner coalitions*, Multi-Agent Systems Engineering, Proceedings of MAAMAW99, Garijo F. and Bosman M. editors, LNAI1647, Springer Verlag, pages 175-187.
9. David N., Sichman J.S. and Coelho H. (1999), *On being responsible: how to be individualistic and smile benevolence to the others*, Proceedings of the 3rd European Conference on Cognitive Science, ECCS99, 27-30 October, Sienna, Italy, pages 57-62.
10. d'Inverno M., Luck M. and Wooldridge M. (1997), *Cooperation structures*, Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI97, pages 600-605.
11. Epstein J.M. and Axtell R. (1996), *Growing Artificial Societies: Social Science from the Bottom Up,* MIT press.
12. FIPA (1997,1998), *Specification part 2: Agent Communication Language,* http://www.fipa.org.
13. Fischer K. and Muller J.P. (1996), *A decision-theoretic model for cooperative transportation scheduling*, Agents Breaking Away, Proceedings of MAAMAW96, Eindhoven, LNAI1038, Springer Verlag.
14. Gilbert N., Sichman J.S. and Conte R. (1998), Multi-Agent Systems and Agent-Based Simulation, Proceedings of MABS98, LNAI1534, Springer Verlag.
15. Hales D. (1998), *Stereotyping, Groups and Cultural Evolution: A case of Second Order Emergence?*, Multi-Agent Systems and Agent-Based Simulation, Pro-

ceedings of MABS98, Gilbert N., Sichman J.S. and Conte R. editors, LNAI1534, Springer Verlag, pages 140-154.
16. Jennings N. (1999), *Agent Oriented Software Engineering*, Multi-Agent Systems Engineering, Proceedings of MAAMAW99, Garijo F. and Bosman M. editors, LNAI1647, Springer Verlag, pages 1-7.
17. Jennings N. and Campos J. (1997), *Towards a social level characterisation of socially responsible agents*, IEE Proceedings of Software Engineering, 144 (1), pages 11-25.
18. Marini M.M. (1992), *The role of models of purposive action in sociology*, Rational Choice Theory - Advocacy and Critique, Coleman J. and Fararo T. editors, Sage Publications, Key Issues in Sociological Theory, pages 21-44.
19. Newell A. (1982), *The Knowledge Level*, Artificial Intelligence, (18), pages 87-127.
20. Prudkov P. and Rodina O. (1999), *Synthesis of Purposeful Processes*, Target Article at PSYCHOLOQUY- American Psychological Association (APA), ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1999.volume.10/, psycoloquy.99.10.070.purposeful-processes.1.prudkov.
21. Rao A.S. and Georgeff M.P. (1991), *Modeling Rational agents within the BDI Architecture*, Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning, Morgan Kaufmann.
22. Russel S. and Norvig P. (1995), Artificial Intelligence, a Modern Approach, Prentice Hall, Chapter 16, pages 473-484.
23. Sichman J.S. (1996), *On Achievable Goals and Feasible Plans in Open Multi-Agent Systems*, Proceedings of the 2nd International Conference on Multi-Agent Systems, ICMAS96, Kyoto, Japan.
24. Sichman J.S. (1998), *DEPINT: Dependence-based coalition formation in an open multi-agent scenario*, Journal of Artificial Societies and Social Simulation, 1 (2), http://www.soc.survey.ac.uk/JASSS/1/2/3.html.
25. Sichman J.S. and Demazeau Y. (1995), Exploiting social reasoning to deal with agency level inconsistency, Proceedings of the 1st International Conference of Multi-Agent Systems, ICMAS95, AAAI Press, pages 352-359.
26. Sichman J.S., Conte R., Demazeau Y. and Castelfranchi C. (1994), *A social reasoning mechanism based on dependence networks*, Proceedings of the European Conference on Artificial Intelligence, ECAI94, pages 188-192.
27. Simon, H. (1957), *A behavioural model of rational choice*, Quarterly Journal of Economics, (69), pages 99-118.
28. Terna P. (1998), *Simulation Tools for Social Scientists: Building Agent Based Models with SWARM*, Journal of Artificial Societies and Social Simulation, 1(2), http://www.soc.survey.ac.uk/JASSS/1/2/4.html.
29. Zlotkin G. and Rosenchein J. (1994), *Coalition, cryptograpphy and stability: Mechanisms for coalition formation in task oriented domains*, Proceedings of the American National Conference on Artificial Intelligence, AAAI94, Seattle, WA, USA, pages 432-437.

Author Index