The Knowledge Engineering Review, Vol. 15:2, 2000, 197–203
Printed in the United Kingdom

Can models of agents be transferred between different areas?*

RUTH AYLETT[1], KERSTIN DAUTENHAHN[2], JIM DORAN[3], MICHAEL LUCK[4], SCOTT MOSS[5] and MOSHE TENNENHOLTZ[6]

[1]*National Industrial Centre for Virtual Environments, Business House, University of Salford, UK*
[2]*Department of Cybernetics, University of Reading, UK*
[3]*Department of Computer Science, University of Essex, UK*
[4]*Department of Computer Science, University of Warwick, UK*
[5]*Centre for Policy Modelling, Manchester Metropolitan University, UK*
[6]*William Davidson Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, Israel*

1 Introduction

One of the main reasons for the sustained activity and interest in the field of agent-based systems, apart from the obvious recognition of its value as a natural and intuitive way of understanding the world, is its reach into very many different and distinct fields of investigation. Indeed, the notions of agents and multi-agent systems are relevant to fields ranging from economics to robotics, in contributing to the foundations of the field, being influenced by ongoing research, and in providing many domains of application. While these various disciplines constitute a rich and diverse environment for agent research, the way in which they may have been linked by it is a much less considered issue. The purpose of this panel was to examine just this concern, in the relationships between different areas that have resulted from agent research. Informed by the experience of the participants in the areas of robotics, social simulation, economics, computer science and artificial intelligence, the discussion was lively and sometimes heated.

2 Embodiment

Does embodiment matter in agent research? If it does, how?

Agent *embodiment* is regarded by Doran as referring, in effect, to the level of complexity, unpredictability, unrepeatability and error in the processes of sensing and action available to an agent. If the level of embodiment in this sense is relatively high, and in particular, if it goes beyond that naturally captured in the agent's internal representations, then the agent is *embodied*. Embodiment is very important, for in its presence the chance of a mismatch between an agent's representations and the reality is high and the agent's task of correctly linking action to perception becomes much harder.

In robotics research, the importance of embodiment is both clear and very visible. Trivially, what a robot can do is determined by its bodily form. For example, no legs implies no walking, and no x sensor implies no perception of x. However, it matters for more subtle reasons too. If a robot is driven by emergent behaviour, then exactly what emerges depends upon the whole interaction between robot and environment; not just the software, but how much the wheels slip, where the sensors are on the body, how well-charged the robot is, and many other things.

In particular, the behaviour of groups of agents may be profoundly affected by the fact that while they may all run the same software, everything else about them is slightly different. No two sensors perform in exactly the same way, no two sets of wheels slip by the same amount, no two robots are equally charged. Where agents are using *stigmergy*—communication via the environment—their differences are a source of variation in the environment for each other. Thus, while it might be

* This report is the result of a panel discussion at the Workshop of the UK Special Interest Group on Multi-Agent Systems (UKMAS'98). All members of the panel are authors, listed alphabetically.

thought that Aylett's two Salford robots, Fred and Ginger (Barnes et al., 1997), might approach an obstacle exactly head-on and try to avoid it by turning in opposite directions, with disastrous results, this has never actually been observed. Small differences mean that one of them is always slightly ahead and *pulls* the other around the obstacle.

Pragmatically, (pseudo) embodiment matters in order to use body language as an additional channel of communication. Thus various virtual agents use gesture, facial expression and stance to communicate emotional state or attentiveness (for example, direction of glance indicates attentional focus). Doing this in robotics is more problematic, because displaying body language with a typical robot body is hard, and noticing it with typical robot sensors is even harder. However, it is quicker to assume cooperation if an agent moves towards an object to be transported than it is to explicitly negotiate. Humans use inference based on body movement at least as much as explicit communication when it comes to physical cooperation.

Is embodiment a specific property of natural or physical agents (e.g., animals, robots), or is it a universal property (e.g., of agent-environment couplings) that can be formalised beyond the concrete physical or software implementation?

If we start from the premise that agency arises from biology, living things have agency but rocks and stars do not. Thus, in a sense, agent research transfers intuitions from nature. Indeed, much agent research assumes a model of autonomous reflective reasoning and human-based mind-body split that is sufficiently abstract for logic rather than biological process to be its target representation.

Conversely, the behavioural approach to robotics, and the whole enterprise of artificial life, might be said to reflect a model of agency in which coupling to the environment and emergent behaviour are seen as the important features and ants are just as much agents as humans. Taking this environmental coupling as a starting point, process rather than object then becomes the organising principle, and a body (as well as mind) can be seen as the temporal propagation of a pattern rather than something static. However, if everything is process, abstracting mind away from body is incongruous with the necessary entwining of processes of body and mind.

Additionally, if emergent behaviour is the focus, then bodily interaction is the source of this emergence rather than a set of software instructions directly mapping to body movements. Indeed, mind may itself be an emergent behaviour. We think intuitively of body as something solid—a physical boundary with an autonomous inside and other agents and processes not controlled by the agent outside. This notion of a static body is rather misleading, however, and clearly so in the case of biological agents; not one cell in an adult human body was present when the human was born and only a continuing process maintains it as a body. It therefore seems preferable to regard a body as a set of processes, but only if the claim of any multi-threaded piece of software to thereby have a body can be excluded. In this process sense, robot *bodies* are extremely primitive and disconnected—while they are full of processes (such as oxidation and electric current), many of these are independent of the software running on them. Robots have no immune system and a very impoverished coupling to the real world when compared to even very simple living organisms.

Distinct from the concept of physical boundary, the ability to move is also important. Bodies consist of elements that can both be moved under autonomous control and support movement of the agent in its environment. This is also somewhat problematic, if only because while we may not think of plants as having bodies in the same sense as animals, it is hard to deny that they are in some sense embodied. Do we necessarily accept that mobile agents (in the Internet sense) must have bodies because they are mobile? Do we accept that an intelligent process plant, constructed as an agent, cannot be embodied because it does not move around?

Even if we take embodiment to be something that actually happens only in the real physical world, we might still see some agent work as concerned with modelling embodiment, and some not. Many software agent systems seem to perform little perception apart from receiving messages from other agents and no actuation to change the environment. They communicate at a symbolic level with others rather than observing them; they agree to carry out tasks but seem to get little feedback

https://doi.org/10.1017/S0269888900002034 Published online by Cambridge University Press

from doing so. It is not always clear what their execution mechanism actually is, and the focus is on deciding to act rather than on acting. This could be problematic if it allows us to produce multi-agent systems without considering the dynamics of their activity (as it might with multi-agent financial systems).

There is also an intuitive difference between a *virtual agent*, with a modelled body and virtual sensors and actuators, and a software agent working at a purely symbolic level. Several virtual agent researchers specifically study gesture, body language and facial expression as a means of communication. Indeed, embodiment concerns some aspect of an agent that is continuously observable by other agents from the outside. This supports inferences about the internal state of an agent and assumptions about its capabilities as well as predictions about its immediate behaviour. It also supports many kinds of inter-agent interaction: a body can allow an agent to be recognised as 'one of us', 'my friend and colleague', 'something good to eat', or 'something that might eat me'.

The question of whether embodiment is a universal property is best answered with reference to the definition provided earlier. In that sense, it is a potentially universal property in that an agent in any domain or of any type may be *embodied*, but embodiment is likely to be a more important issue in some domains (such as robotics) than others. Indeed, if embodiment is physically contingent then it may have to be located within concrete physical systems if one is to say anything interesting about it. Models of embodiment may be interesting research tools and may help us to build more useful agent systems, but may not themselves be embodiment. Virtual rain does not get anyone wet.

Do you think that agent models and theories can be transferred between different kinds of agents like virtual, robotic, software, or animal agents (e.g., ethological models on animal behaviour)? Do you know examples of such attempts? How do you see the applicability and limitations of such an approach in your specific area of agent research?

Here, the panelists offered a range of responses. Doran argues that the range of possible core structures of agents (that is, the *architecture* of the agent within its embodiment, if any) are common over all agent (application) domains (Doran, 1996). For example, there is no substantial difference in the core agent structures that seem to be available in mobile robotics, intelligent buildings, social simulation, and network management, although terminology is to some extent divergent.

Similarly, Aylett suggests that as above, everything in this field is based on varying analogies from nature. Much interesting work in robotics tries to reapply what is known about ants, bees, crickets, slugs, cockroaches and other insects. Aylett's work on cooperative robots at Salford started from social insects and not from humans, looking for ways in which cooperation could be produced as an emergent behaviour from physical interaction rather than as an explicit result of reasoning. Here, cooperation was *bodily* activity rather than a *mental* one.

There is no reason why the same ideas and models cannot be reapplied to virtual agents (specifically agents in virtual environment) and to other software agents. Much artificial life work uses simulation as, for example, the Santa Fe Trail (a well-known ant model) and prey-predator simulations, to explore models based on the natural world. In particular, much work in synthetic or virtual agents reuses robotic models, as in the application of the Salford behavioural robot architecture to virtual Teletubbies (Aylett et al., 1999), for example. Similarly, the JACK project consciously reapplies a typical three-layer robotic architecture to the animation of a biomechanical model of the human body. Many other robotic architectures (such as Firby's RAP) have also been reapplied. It is easier to reapply robot models to virtual agents—as compared to other software agents—since the environments in which they are located are themselves (fairly crude) models of the real world in most cases. In fact, because of the limitations of engineering and cost, it is actually possible to model more sophisticated agent-environment couplings in virtual worlds than robots achieve in the real one—at least from the agent side. For example, the system of chemical emitters and receptors used in the games software Creatures is well ahead of any robot implementation.

The down side of this is that while it is easier to model the agent, it is harder to deal with the environment to which it is coupled. This is because in the real-world physics just *is*, while in a virtual world, physics must be added explicitly. Virtual worlds have a tendency not only to be impoverished but to be uniform, because this is the easiest thing to do. If robotics often provides rudimentary agents in an interesting world, virtual agents can be seen as providing interesting agents in a rudimentary world.

A more reserved assessment was made by Moss, who recast the original question to ask whether and under what conditions such a representation of agents and agent cognition can be transferred from social simulation models to useful systems in robotics, information search and process control. He illustrates the discussion with the example of his work with Schroeder at City University to develop specifications of software agents so that they can autonomously develop trading procedures and practices as well as contacts for buying from and selling to other information agents. The idea is that trading arrangements, practices, norms etc.—what most of us would call *markets*—will emerge from the activities of the software agents. The emergence is important because we know that, in the real world, the characteristics of markets (i.e., the trading arrangements, norms, etc.) differ for different commodities. Ships are simply not sold in the same way that cars or chocolate bars or even boats are sold. Arguably, the differences in the ways that different commodities are traded depends upon their physical characteristics and the technology involved in effecting transactions.

Now some people working in MAS and, in particular, on information agents, simply adopt approaches to agents and markets taken from economics. This is a curious approach since economists do not themselves argue that their representations of agents are descriptively accurate and they do not analyse the process of exchange. The more promising approach is to start with historically accurate descriptions of the emergence of various trading practices, norms, etc., taking account of the relevant technologies involved. Then we can reason about the ways in which the historical experience can apply to information markets where autonomous software agents trade. That process of reasoning is suitably conducted within the framework of social simulation modelling.

As a second example, Moss describes work with Dautenhahn on the use of organizational structures to control robot teams engaged in certain types of activities (Moss and Dautenhahn, 1998). The control mechanisms are specified on the basis of models of real organisations though they are by no means straightforward copies. The differences in the models of organisations of real people and the organisation of simulated robot teams result from reasoning about the differences in the problem domains. A further step would be to test the simulations on actual robots.

In a similar vein, Tennenholtz argues that while basic ideas from one area can be borrowed by another, as was demonstrated by his work on artificial social systems (and its application to robotics) and on protocols for non-cooperative agents (and its application to electronic commerce), in many cases, these ideas may lead to approaches that are vastly different from the ones used in the original domain. For example, social laws in robotics are usually quite different from the ones used in human activity, while models of protocols for non-cooperative agents applied to electronic commerce deviate from the classical economic assumptions from which they were derived.

3 Interdisciplinarity

Is it possible and/or desirable, and/or necessary to transfer models and theories between different areas, like social science, anthropology, MAS, robotics, etc? How do you see differences in goals and objectives of agent research in these different fields? (Are we talking the same language?)

Human sciences can tell us much about the constraints under which humans act, and it seems silly to ignore this knowledge. Anthropology seems particularly relevant in that it looks at the links between the behaviour of individuals and the behaviour of the whole society for a variety of different cultures. It is also extremely important—to the point of necessity—to be aware of biological models and concepts and how they can be reapplied in computer-based agents of various kinds. Complexity, the current state of biological understanding, and computational resource seem to be some of the main problems in doing this.

https://doi.org/10.1017/S0269888900002034 Published online by Cambridge University Press

Other disciplines that should be mentioned are psychology and the arts. Psychology is another field from which useful models can be garnered, with the same proviso as with social sciences, that psychological models sometimes have a rather tenuous link with empirical psychology. Work in synthetic agents is heavily influenced by work in psychology on emotion and behavioural interaction, for example. Artistic theory is also being reapplied, particularly from the theatre and drama, where issues of role playing, improvisation and believability, as well as narrative coherence have all been taken up in work on synthetic, or virtual, agents.

In all these cases, there is every chance that we are all speaking different languages—using different terms to mean similar things or the same terms to mean quite different things. This seems to be all part of the fun, however, and will resolve itself as long as people keep talking to each other.

Clearly, transferring models and theories is both possible and desirable. However, as Doran points out, *pure* theories of core agent architecture and of multi-agent systems (analogous to pure mathematics) can be distinguished from their application in some particular application domain (analogous to a piece of pure mathematics finding an application). In considering the applicability of models and theories in different domains, therefore, we can distinguish between attempts to develop new pure theories of agent systems, attempts to engineer effective agent systems for particular tasks such as network management, and attempts to use agent systems to model social systems, for example, to learn about the latter. The relevance from each of these concerns may be more or less applicable in other areas.

Indeed, theories cannot always be simply transformed from one area to another area; they must be carefully reconstructed, and require significant familiarity with related theories. For example, many problems in electronic commerce are some mixture of economics and computer science questions, but require a deep understanding of related theories before they can be effectively modified. Finding general solutions that can be transferred across domains may not be a simple process.

In transferring models, it is important to be aware of the different levels on which they work. Robots and virtual agents must cope with physical movement in a spatial world. The problems and issues faced here may operate at a different level to macro social science, where it is still controversial how complex the model of an individual agent really needs to be. One should also be aware of what assumptions are encoded in a social science model—for example, economic models frequently assume rational individualism. Finally, one should be aware that social science models are not always firmly connected to empirical social science—it is for that very reason that the use of computing can help social science theorists, since it animates the models and demonstrates their empirical consequences.

According to Moss, this kind of experience indicates clearly that models of agents *can* be transferred among the various MAS disciplines. Whether it is useful depends upon the models being transferred and why. The fundamental issue seems to be less about transferring models than transferring the understanding and experience that led to the formulation of those models. If we have learned about markets in the real world and why different markets entail different arrangements for determining supplies, prices, demands and the characteristics of what is being traded, then that understanding should help us to reason about the characteristics of systems that will allow information, or other, markets to emerge.

We can also learn about analogous relationships in (say) information markets or robot control systems, etc., and use that additional understanding in our analyses of social systems. Moss believes that, in all such cases, good science involves clear specifications of the conditions in which a model or representation of agents is applicable and the class of results from using that model or specification that will confirm the appropriateness of that model or specification or, alternatively, will indicate that it is not appropriate.

https://doi.org/10.1017/S0269888900002034 Published online by Cambridge University Press

What are the differences between software agents, e.g., in fields like MAS, economics, and social simulation studies? Will social sciences help us building successful agent-based systems?

The range of core structures of agents is essentially independent of the application domain. As argued above, there are differences of granularity in some of these fields; the internal complexity of agents in a multi-agent system may be much higher than that in a social simulation. On the other hand, the types of interaction between agents may be much more varied in a social simulation than they are in a multi-agent system—in the former, explicit communication may be supplemented by all kinds of indirect communication via the state of, and changes in, the simulated society.

Economic models may have the least varied styles of interaction between agents since they usually assume that most information is carried by global prices and their movements rather than by explicit communication. Both agent systems and economic models may assume rational models of agenthood; social simulations are less likely to do so. It should, however, be noted that in the case of economic models, this depends upon whether the aim of the model is normative (to show how an *ideal* market should work) or descriptive (to show how a real market does work).

Tennenholtz further suggests that experimental studies in the social sciences are crucial for some of the work on agent technology such as electronic commerce, especially when there are human users. In this case, experiments by cognitive psychologists and experimental economists in order to test whether a protocol or algorithm is useful, can be vital. However, it seems at least as plausible to ask whether computer modelling will help to build successful social theories as to ask whether social science will help to build successful agent systems. Some expect the engineering (rather than modelling) applications of multi-agent systems to continue to be the main driver of agent and multi-agent theory over at least the next decade, but no discipline has a monopoly on either good or bad ideas.

What are the general advantages and limitations of such attempts? How do you see the applicability and limitations of the attempt to import models from other areas in your particular area of agent research?

As far as *social laws* for robots are concerned, Tennenholtz agues that we are now in a situation where we have extracted from the social world a design paradigm which, once formalised, becomes almost a pure engineering task. On the other hand, work on protocols for non-cooperative agents still requires effort to bridge the gap between computer science theories and economic theories—we cannot bypass these gaps in a simple way. In general, these attempts are very useful; the transition from logical theories to decision-theoretic models in mainstream artificial intelligence is one useful example of this (which is not to say that it is a better approach, of course) that brings new ideas and application. Such a transition could not happen if we had not tried to create these bridges between theories in different areas.

Moss responds from a different perspective. Specifications of agents, their cognitive capacities and the ways in which they interact are usefully informed by simulations conducted for social science purposes. Moreover, the application of social simulation models to data mining, robotics and process control provides a test of social theories and modelling procedures. Indeed, his own experience suggests and his research programme is based upon the proposition that reasoning about the transfer of agents from one discipline to others can usefully inform the *receiving disciplines* and provide tests of the specifications of the originating disciplines.

Any discipline can get bogged down in its background assumptions—those ideas that are so universally held that nobody is conscious of them any more. The advantage of inter-disciplinarity is that different disciplines have different background assumptions, so that interaction can make everyone more aware of what they are assuming.

References

Aylett, RS, Horrobin, A, O'Hare, JJ, Osman, A and Polyak, M, 1999, "Virtual teletubbies: reapplying a robot architecture to virtual agents" *Proceedings Third International Conference on Autonomous Agents*, pp 338–339. ACM Press.

Barnes, DP, Ghanea-Hercock, RA, Aylett, RS and Coddington, AM, 1997, "Many hands make light work? An investigation into behaviourally controlled co-operant autonomous mobile robots" In: Johnson, WL and Hayes-Roth, B (eds.), *Proceedings First International Conference on Autonomous Agents*, pp 413–420. ACM Press.

Doran, J, 1996, "Simulating societies using distributed artificial intelligence" In: Troitzsch, KG, Mueller, U, Gilbert, GN and Doran, JE (eds.), *Social Science Microsimulation*, pp 381–393. Springer-Verlag.

Moss, S and Dautenhahn, K, 1998, "Hierarchical organisation of robots: a social simulation study" *Proceedings 12th European Simulation Multiconference ESM98*, pp 400–404. Manchester, UK.