Agent-Based Social Simulation *and its necessity for understanding socially embedded phenomena*

Bruce Edmonds

Abstract

Some issues and varieties of computational and other approaches to understanding socially embedded phenomena are discussed. It is argued that of all the approaches currently available, only agent-based simulation holds out the prospect for adequately representing and understanding phenomena such as social norms.

Cognitive Simulation Modelling

For the last few decades computers have been used to model cognitive processes (e.g. Newell and Simon 1972). That is computer programs are made that allow the simulation of aspects of human cognition. This field has grown over the years in parallel to that of artificial intelligence, which is different because it aims to implement aspects of intelligence using computer programs but not necessarily in the way humans achieve this.

Cognitive modelling comes in a number of different purposes, with different levels of realism and pursuing a variety of different goals. However, at least within the field, the usefulness of cognitive modelling is well established, for even if a particular model or simulation turns out to be mistaken (i.e. for the model purpose the brain turns out to works in a significantly different way) having to instantiate an idea about the workings of our cognition forces the model to be: (1) complete (no hidden explanatory gaps) (2) explicit (no vagueness) and (3) feasible (has to be able to be computed within a reasonable amount of time. Thus instantiating a theory in a computation model constrains theory in useful ways. Of course, if the model can be constrained by evidence of how human cognition happens to work, that is even better.

Although there have been many architectures and frameworks for cognitive modelling SOAR and ACT-R have attracted the most researchers. Each of these has evolved to become substantial sub-fields, encompassing a whole host of models. However these are not ideal for capturing social aspects of cognition because: (1) they are quite computationally heavy thus making it difficult to include the interaction of many agents (Ye and Carley 1995) included 3 interacting SOAR agents but it required a separate computer running each one) (2) their input/output facilities are not so well supported and (3) they are overly complex for most social simulation purposes, for example in the synchronisation of agent actions which would require explicit signalling in SOAR/ACT-R. However the main reason is that the researchers involved have been focused on individual cognition and not very concerned with the social interaction of agents, maybe assuming that this is something to be dealt with *after* having sorted out the cognitive model.

More recently agent-technologies, such as BDI (Belief, Desire, Intention) that are specifically inspired by human cognition[1] have been developed, supported by a logic-based approach (Rao and Georgeff 1998) which allows for reasoning about beliefs, desires and intentions to be done by software agents.

[1] In the case of BDI (Bratman 1999)

Agent-Based Architectures and Frameworks

More recently the field of "Software Agents" or "Multi-Agent Systems" has developed its own series of architectures. These can be broadly classified as "cognitive" but the connection between human and agent cognition is very much looser. As in AI there is no necessity that software agents work in the same way as humans do. However there are several reasons why human cognition, and in particular human social cognition remains the primary source for ideas as to the necessary structure and processes in agent cognition.

Firstly, effective cognition (that is cognitive structures and processes that allow an agent to operate within its environment in an autonomous manner) is difficult to arrange but is obviously something humans manage to the degree they do. Thus systems inspired by or derived from how humans think are a rich source of ideas for how to endow software agents with commensurate abilities. Thus abilities such as: reasoning, sub-dividing problems, pattern-recognition, associative memory etc. are all sources for implemented and tested agent processes.

Secondly, the essentially social problems that an effective agent has to deal with have a lot in common with those humans cope with. Thus issues such as social recognition, trust, reputation, obligation, negotiation, communication, speech acts etc. are all ideas that have a direct application in multi-agent systems. Thus concepts such as trust and obligation have been formalised as part of a framework for understanding what these might mean in the extra-human context of software agents (e.g. Conte and Castelfranchi 1995) and ideas taken from social science have been explicitly applied within distributed computational systems (e.g. Hales and Edmonds 2005).

Thirdly, it has been discovered that specifying and designing effective multi-agent systems can be facilitated by an analysis based on social-roles (Wooldridge et al 2000). Thus there are a number of methodologies that use a quasi-social analysis, which identifies roles which agents might fill which are defined in terms of the rights, obligations, protocols etc. which pertain to that role as an aid to the specification of a multi-agent system .

As in cognitive modelling, what could *work* in a social setting does provide *a priori* constraints upon the possible theories and architectures that might lie behind social norms, clearly in order to understand how *human* norms might work this is insufficient. However due to the close parallels between the sort of processes used in multi-agent systems and those thought to occur in human social systems, the techniques and technology of multi-agent systems make the ideal tool for analysing the complex and intertwined processes involved in social norms.

The Social Intelligence Hypothesis

The Social Intelligence Hypothesis (Kummer et al 1997) says that the evolutionary advantage of human intelligence (and to a lesser degree the intelligence of the great apes) lies in the ability to relate in socially sophisticated ways. These social abilities allow humans to cooperate, form and maintain groups, communicate, teach information to the next generation, know who to trust, gossip etc. Together these abilities allow groups of humans to survive in a variety of niches (Tundra, Kalahari desert etc.) where individual humans (even very clever individual humans) could not. They seem to achieve this by the development, maintenance and adaption of group cultures of technologies and social institutions that allow survival in each niche (Reader 1988). Thus this hypothesis combines a plausible theory for the evolutionary advantage of our intelligence as well as explaining many of its unique characteristics.

If this hypothesis is true then the social abilities of humans are not merely an "add-on" to our general intelligence nor an outcome of an otherwise evolved intelligence, but are the core and reason for our intelligence (Edmonds and Dautenhahn 1998). Rather it is our "general" intellectual abilities that are the "by-products" of our social intelligence – for example the fundamental utility of language is for communication and speech acts but it also happens to be useful for externalising and formalising reasoning. To understand our intelligence requires understanding its social abilities, abilities that will only make sense in a social context. However, understanding how our abilities work *within their social context* is very difficult. If one simply observes what people are doing in a social context one can not see the corresponding changes in the cognition of the actors involved; if one carefully determines the cognitive processes in laboratory experiments one misses most of the social context in which the abilities make sense.

Social Embeddedness

Grannovetter (1985) pointed out that a significant portion of human behaviour is socially embedded – that is to say, it can not be properly understood in terms of either under- or over-socialised models. The under-socialised model is when an individual's behaviour is considered entirely in terms of its self-interest and rationality, with the surrounding society reduced to the 'environment' against which this rationality reacts. The over-socialised model assumes that the dictates of society are so internalised and widespread that the individual can be forgotten and sufficient understanding derived from studying how people, en masse, behave using statistics and the like. In other words, social embedding implies that there are many aspects of human life that can not be satisfactorily modelled if one omits either the: individual and its decision making or the system of groups and interactions that the individual is embedded in. In particular one will miss key phenomena if one tries to reduce such situations to some mass social trend or if one models the individual as a single entity interacting with a token environment or problem. In other words, the detailed patterns of how individuals interact with other individuals matter.

If the Social Intelligence Hypothesis is true, and our intelligence is of a fundamentally social nature, then it will be almost a hallmark of our behaviour that it brings about, and indeed depends on, intermediate social structures. The SIH implies that these social structures have vital survival value and thus can not be ignored. Thus the SIH implies that many crucial aspects of human action can only be understood from a socially embedded viewpoint – the coordination and institutions are not merely an afterthought or epiphenomenon, but the *essence* of why much human action is effective.

If considerable portions of human behaviour is socially embedded in this above sense, then its understanding needs more than a good cognitive model, more than a cognitive model that includes social abilities, but a model which also captures the patterns of individual interactions between individuals. Thus the social networks and the interactions within these networks matter in important ways, ways which mass statistical methods will simply miss.

Micro-Macro Complexity

The prevalence of social embeddedness causes a problem for those who wish to understand human social behaviour. The link between the individual behaviour and the characteristics observable at the aggregate level is very complex – one can not simply average out the individual actions to get to the population characteristics, since the intricate and intermediate patterns of social interaction make this difficult. This is the micro-macro problem of sociology: how to relate the micro behaviour and characteristics of individuals to the macroscopic characteristics and trends as measured by surveys, polls, and other social statistics.

If one is trying to understand situations where there is a high level of social embedding, then a project of understanding social trends and affects without dealing with some of the intermediate level (i.e. local social level) detail is doomed to superficiality. These structures, processes and institutions mediate the effect of individual actions and determine how such actions aggregate up to the societal level.

Where the aggregate level seems to behave in a way that is "qualitatively beyond" that of its components (the social actors), this is called an emergent phenomena. This is not the place for a discussion on what “qualitatively beyond” might mean, but roughly it implies that the most appropriate language of description is different at macro and micro levels. Society is full of emergent phenomena, for example fashions, or the booms and busts of the stock market. However, unlike physical, chemical or most biological phenomena, social phenomena not only involves an upward emergence of phenomena but macro scale phenomena have a “downward” effect on the individuals’ cognition (sometimes called immergence). Thus it is crucial to the phenomena of fashion that not only do people tend to follow a certain fashion in increasing numbers, but that the fashion is recognised by the individuals concerned which will affect their behaviour and the subsequent emergence of trends in clothing – for example once a fashion has been recognised and labelled this might have the effect of locking-in some followers to subsequent changes of trend *within* the confines of the labelled fashion and others to reject it, because it is part of a particular fashion when otherwise they might have adopted it.

Once one has both emergence and immergence then this loop can itself result in different kinds of interactions and societal patterns, resulting in yet more complexity. It is this “double” complexity that characterises a lot of the social world.

Types of Social Simulation

In all types of social simulation there is a computer program that constitutes the detailed specification of the model and the (implicit or explicit) mapping from the states of the computer that result to what is represents. It can happen that the same program can be used as a model for different targets with different mappings. For example it has been quite common for some simulations with an evolutionary “flavour” that they could be interpreted in two ways: in a biological manner with the reproduction of genes or in a social manner with imitation of some behaviour.

Simulations can have many different purposes. Epstein (2008) lists 17 purposes. Four key purposes are: prediction, explanation, exploration (of the nature of plausible processes), and illustration (of ideas). A predictive simulation will accurately anticipate some aspect of an unknown situation for the type of system or situation modelled, though this does not necessarily mean a point prediction of a value, but could include weaker versions of prediction, such as that a certain results will not occur, or that a certain kind of outcome will result (e.g. Models of Darwinian evolution predict that there will be a tree-like development of species but does not predict which species will develop). An explanatory simulation may well be fitted to available data, in which case it provides a computation from the setup in terms of the processes specified for the simulation, which (interpreted via the mapping to the target of modelling) provides a candidate explanation for those outcomes in terms of those processes. An exploratory simulation may not have any reliable mapping to what is observed, but is used to discover some of the possible complex outcomes that result from the interplay of the specified processes, this can be used to produce examples (or, more powerfully) counter examples and establishes a basic computational plausibility (but only as so far as the specified processes and settings are plausible). The purpose of an illustration is to demonstrate an idea, it needs to be as clear as possible, but beyond a kind of plausibility that helps us relate to it, has no obligation to correspond to anything observed, but rather an idea or theory.

Different purposes imply different criteria for the usefulness of the simulation and also for the *way* that a simulation is developed for this purpose. For example if the purpose is explanation then *how* the simulation mechanisms are specified, in particular how closely they relate to the mechanisms which explain the outcomes, is important otherwise the explanation one gets is not relevant for the case under study. If one was using a simulation for the purpose of prediction, then it would not matter so much *how* the results were obtained but more the accuracy of the outcomes in the relevant aspects.

Another way in which simulations differ is in their level of detail, or their degree of aggregation. Statistical and system dynamics models might represent the whole of society as essentially one object which has different related properties. This is a minimal level of aggregation, with the noise representing the imperfect nature of this representation. However, as discussed above, social embeddedness means that this will miss out many key aspects of some social phenomena. Given the socio-cognitive nature of norms and the importance of social structure to these (e.g. social networks and groupings) a more detailed representation is likely to be required.

An individual-based simulation, represents each social actor as a separate entity in the simulation – each with its own properties. This allows for the complex interaction of the actors to be explicitly represented. If the representation of the individual has elements of cognition (e.g. individual reasoning or learning processes) then the simulation is called an agent-based simulation. Obviously what counts as an agent depends somewhat on the interpretation of the individuals in the simulation. This can be compared to Dennett's "Intentional Stance", which is considering the object as having intentions because this is useful to do so[2]. One *can* consider all sorts of things *as if* they had intentions, for example a that a thermostat has an intention to keep a room at a particular temperature, but *how* useful this is can be a matter of debate. Similarly the individuals in many simulations can easily be considered as agents since their actions (and hence decision making processes) are mapped to those of social actors, which *do* have cognitive processes. However in many cases although the individuals may *represent* a social actor with cognition, the *representation* may be so simple and lacking in internal processes that calling the individual in the simulation an agent might make little sense.

This highlights the distinction between what exists in the simulation itself, what exists in the phenomena being modelled, what exists in the interpretation of the simulation and (finally) what exists in the ideas that were behind the simulation design. Although frequently conflated in many accounts these can be very different. Many simulations do not attempt to relate the contents of a simulation with that of what is observed, rather they model a (explicit or otherwise) theory or idea about the phenomena of concern – what the researcher *conceives of* as occurring. The theory or ideas then relate to a greater or lesser degree to the evidence (Edmonds 2001). Unfortunately researchers often conflate their conception of what is happening with the phenomenon itself, seeing the world through their "theoretical spectacles" (Kuhn 1969).

Thus a key dimension along which simulations vary is its "distance" from the evidence. Purely theoretical simulations will concentrate on the properties of the model itself, with only the vaguest of motivations from any evidence. Some of these abstract explorations have the flavour of a computational analogy – more of an illustration of an idea with which to think about some phenomena than a articulated and applicable theory in itself. Evidence-led or descriptive simulations will be specified without the help of much grand theory, but seeks to capture the available evidence into the coherent and dynamic framework of a single simulation. Of course most simulations hover somewhere between and it is often unclear the extent to which they relate to the evidence, and might only indirectly do so.

Linking Plausible Theory and Observed Evidence

For the reasons discussed above, understanding social processes and phenomena is hard. Any usefully-identified causation between the micro-abilities and propensities of actors and the global outcomes in the society or group will be intermediated and constrained by the transient social structures, norms, agreements, fashions etc. that emerge and dissolve. These rich and dynamic social structures make it difficult to understand the phenomena using only methods of: social statistics, social psychology or abstract economics that, in effect, represents society using a single entity.

[2] Of course, if it *is* very useful to consider that something has intentions this suggests that it *may* actually have intentions or something like them. If it looks like a duck, smells like a duck, tastes like a duck.....

One way of attempting to bridge the gap is via an intermediate, "bridging" concept. Examples of these ideas include "social capital" or "norms" which link the individual behaviours of actors to the observed social outcomes. These are explanatory ideas, something which is not necessarily directly observable but allow some understanding of the causal links involved. However since these are necessarily discursive ideas they inevitably lead to a variety of interpretations. Their generality allows them to cover the intermediate social complexity, but this also leads to debates about their definition when they are used in earnest to explain specific cases. In a sense we are asking too much of a discursively-defined idea – even if an idea was right[3] the imprecision of its definition would mean that we could not tell what its ramifications were in enough detail to know if it did pertain or not in many particular cases.

Another approach is to pass over the intermediate complexity between micro and macro levels, and to attempt to discover connections between cause and effect at the macro level by statistical means. This can detect trends and correlations in the broadest sense, but has to ignore the effects of social embedding and any intermediate and/or transient structures the individual is involved in. The difficulty in this approach is the degree to which the individual deviations from these trends can be considered as noise – irrelevant to the phenomena of concern. There are simulations that show that even intermediate transient social structures can have a significant effect on the global outcomes[4] - in other words it is not only the global pressures that matter. Thus, whilst a broad examination of correlations at the macro level can give useful insights, one can not assume it is sufficient.

Analytic models, in the sense of formal equations whose outcomes are obtained by solving them, are capable of capturing some complex processes, and allowing complicated outcomes to be tracked through proof. However, their complexity is severely limited due to the requirement that they be soluble. Of course it is always possible to numerically calculate their consequences for particular initial conditions, but this is merely a kind of simulation anyway. It is largely irrelevant by which formal means one programmes a simulation – the continuous, the discrete, the rule-based, the logic-based, the procedural etc. are all able to approximate each other to arbitrary accuracy – what matters is what it represents and how[5].

Thus the only available tool adequate to the understanding of socially embedded phenomena such as social norms is agent-based simulation. It is only by the tracking of the intricate interplay between the cognitive and the social that such processes can be adequately formalised and understood. That is not to say that other kinds of approaches do not have a role to play, simply that, *currently*, for a full and rich understanding of socially embedded phenomena, agent-based simulation is necessary.

Relevance vs. Generality in Simulation

Of course, simulation is not a magic bullet. The fact that one simulates something like the processes involved in social norms does not make the considerable difficulties disappear. An ideal model of social phenomena would be both relevant *and* general. Relevant, in this context, means that the outcomes of the simulation can be strongly related to the micro and macro evidence available, whilst the generality of a model implies it *abstracts* from particular cases which usually means it is less complicated than these cases. The tension between relevance and generality it not one easy to bridge. When dealing with socially embedded phenomena simulations that are relevant are likely to be complicated, whilst abstract simulations are likely to be relatively simple (especially those we

[3] Right in the sense of providing the best possible explanation of how micro and macro levels relate.

[4] For example the high average level of cooperation in one-shot prisoner's dilemma games when there are simple social grouping mechanisms due to the continual waxing and waning of cooperative groups (e.g. Hales 2000 or Aktipis 2004)

[5] Thus it is perfectly possible to specify an agent-based simulation using differential equations, though it is doubtful that there is any advantage in doing so.

have a chance of understanding). This tension is sometimes expressed as a simplicity-complexity trade-off (Edmonds 2005). The sad fact is that we can not assume that we happen to have evolved brains that are capable of fully understanding models that are adequate in terms of their social relevance.

One possible answer is not to rely on a single model, but attempt to build a closely-related cluster, of models (Giere 1990), including some complicated models that can be directly related to the evidence and simple ones that we have a chance of understanding. If the simpler models that we understand can be mapped onto processes and data gained from complex but descriptively adequate simulations then we may be able to get some of the best of generality *and* relevance, albeit at the cost of complexity in terms of the number of models and mappings between them, and the effort required in building, checking and maintaining such clusters of models.

Emergence and Immergence in Simulations

It is now commonplace that simulations can be used to show how the interaction of many non-random but independent entities can result in complex outcomes that seem to go “beyond” the “sum of the parts”, i.e. that demonstrate emergent effects (Gilbert and Troitzsch 2005). However a simulation also allows the constraint of the parts as a result of the macro-level effects, via the mechanisms of social institutions (e.g. laws) or highly correlated pattern recognition by individuals (as occurs in fashions). This downward causation, from the macro to the micro, is sometimes called “Immergence”. Individual-based simulation (including agent-based simulation) allows for both emergent and immergent processes to be represented – the only formal way of doing so. Thus simulation allows for the interplay of these processes to be explored and studied in a way that seems impossible to do otherwise (with sufficient rigour to enable the tracking of the intricate embedded processes involved).

Conclusion

Agent-based simulation, which represents significant aspects of both the cognition of actors – their social interactions and the societal level constructs – is the only feasible way of understanding the tangle of complex social phenomena, such as those that involve norms.

References

Aktipis, C. A. (2004) Know when to walk away: contingent movement and the evolution of cooperation in groups. Journal of Theoretical Biology 231, no. 2: 249-260.

Bratman, M. E. (1999) Intention, Plans, and Practical Reason. Univ. of Chicago Press.

Carley, K. M. & A. Newell (1994), The Nature of the Social Agent. Journal of Mathematical Sociology , 19(4): 221-262.

Conte R. and Castelfranchi C. (1995). Cognitive and social action. London: London University College of London Press.

Edmonds, B. & Dautenhahn, K. (1998) The Contribution of Society to the Construction of Individual Intelligence. Workshop on Socially Situated Intelligence, at SAB'98, Zurich, August 1998. http://cfpm.org/cpmrep42.html

Edmonds, B. (2001) The Use of Models - making MABS actually work. In. Moss, S. and Davidsson, P. (eds.), Multi Agent Based Simulation, Lecture Notes in Artificial Intelligence, 1979:15-32.

Edmonds, B. (2005) Simulation and Complexity - how they can relate. In Feldmann, V. & Mühlfeld, K. (Eds.) Virtual Worlds of Precision - computer-based simulations in the sciences and social sciences. Lit Verlag, 5-32

Epstein, J. M. (2008). Why Model?. Journal of Artificial Societies and Social Simulation 11(4)12. http://jasss.soc.surrey.ac.uk/11/4/12.html

Giere, R. (1990) Explaining science: a cognitive approach. Chicago University Press.

Gilbert, N. & Troitzsch, K. G. (2005) Simulation for the Social Scientist, Open University Press.

Granovetter, M. (1985) Economic Action and Social Structure: the Problem of Embeddedness., American Journal of Sociology, 91:481-93

Hales, D. & Edmonds, B. (2005) Applying a socially-inspired technique (tags) to improve cooperation in P2P Networks, IEEE Transactions in Systems, Man and Cybernetics, 35:385-395.

Hales, D. (2000) Cooperation without Space or Memory: Tags, Groups and the Prisoner's Dilemma. In Moss & Davidsson, (eds.) Multi-Agent-Based Simulation. LNAI 1979:157-166. Springer. Berlin.

Kuhn, T. S. (1969) The Structure of Scientific Revolutions, University of Chicago Press.

Kummer, H., Daston, L., Gigerenzer, G., & Silk, J. (1997). The social intelligence hypothesis. In P. Weingart, S. D. Mitchell, P. J. Richerson, & S. Maasen (Eds.), Human by nature. Between biology and the social sciences (pp. 157-179). Mahwah, NJ: Erlbaum.

Laird, R., Newell, J. & Paul, A. (1987). Soar: An Architecture for General Intelligence. Artificial Intelligence, 33: 1-64.

Wooldridge, M., Jennings, N. R. & Kinny, D. (2000) The Gaia Methodology for Agent-Oriented Analysis and Design. Autonomous Agents and Multi-Agent Systems 3(3): 285-312

Newell, A., & Simon, H.A. (1972). Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.

Rao, A. S. & M. P. Georgeff (1998) Decision. Procedures for BDI Logics, Journal of Logic and. Computation, 8:293-343.

Reader, J, (1988) Man on Earth. Collins.

Ye, M. & K. M. Carley (1995), Radar-Soar: Towards An Artificial Organization Composed of Intelligent Agents, Journal of Mathematical Sociology 20(2-3): 219-246.