

## **When does social simulation need cognitive models?**

**Nigel Gilbert**

Contributors to this volume have explored the ways in which cognitive models or architectures may be helpful or even essential for building simulations. In this epilogue, I shall be considering whether cognitive models are always necessary – is a social simulation necessarily inadequate if it has no or only a very simple model of cognition? If not, is it possible to specify classes of simulations for which cognitive models are necessary or unnecessary?

I begin by rehearsing the issue of ‘levels’, which has been touched on by a number of contributors, suggesting that analytically at least it is possible to distinguish a biological, a cognitive and a social level, in which the characteristics of phenomena at one level are emergent from the behaviour of phenomena at levels below. This leads to a consideration of when social models need to take account of the details of cognitive architectures (and when cognitive architectures need to take account of social phenomena). Finally, I discuss the problem of how to select among the cognitive architectures on offer when it has been decided that one does need to include a cognitive model in one’s social simulation.

The idea of ‘levels’ is quite difficult to pin down, although commonplace not only in the social simulation community but also more generally in science. It expresses the idea that small-scale details can be abstracted away when considering phenomena at a more macroscopic scale (Gilbert, 1995). For example, to understand the behaviour of

ordinary physical objects, you do not need to know about the composition of atoms; similarly to understand ecology, you do not need to be familiar with intra-cellular processes. Just as, conventionally, a distinction is made between sub-atomic and atomic, and between cellular and ecological 'levels', a similar distinction is made between the cognitive (i.e. the individual) and social 'levels' in the social sciences. Social theorists often make a further distinction between the organisational and the social.

Phenomena at a higher level are said to 'emerge' from behaviour at the lower level. Emergent phenomena are ones where there is an observation mechanism for the emergent phenomenon that does not apply to structures at the lower level (Baas & Emmeche, 1997). For example, the mind is an emergent phenomenon of the brain; political parties (and all social organisations) are emergent phenomena of interactions between individuals; and identity is an emergent phenomenon of cognition. In each case, what we can observe at the higher level (a political party, for instance) is not a possibility at the lower level (an individual cannot be a political party). It is important to note that the definition of emergence involves an external 'observation mechanism'. This may be as simple as collecting public information about the existence of a political party, or as complex as magnetic resonance imaging, but without an appropriate observation mechanism, emergence cannot be detected. Different mechanisms can yield different observations and so different conclusions about emergence. This implies that emergence and indeed the separation of structures into levels is a matter of scientific convention: there is a sense in which it is true to say that emergence is in the eye of the beholder.

The reductionist programme of scientific research argues that one should start with immediately perceptible phenomena and then dive down into successively deeper levels. This programme has been outstandingly successful in physics, effective to a degree in biology, where it has given us molecular biology and the human genome programme, and may also be part of the justification for this volume where contributors argue for the importance of understanding the cognitive in order to understand the social. But humans differ from atoms and ants, in that they are able to comprehend macro patterns (what sociologists call institutions) and these institutions can change individual behaviour. For example, voters can be influenced to support a political party by the success of the party's campaign, which was intentionally constructed by its leaders to maximise its support. Here individuals are influencing the party's platform, and the platform is influencing the voter, who might join the party and help to shift its political priorities. The example illustrates the interaction between the organisational and individual levels which, if successful, could become self-reinforcing. Neither a reductionist programme that attempts to explain the politics of the party from the actions of its individual members, nor a structuralist programme which aims to understand the political actions of the individuals solely by reference to the manifesto of the party are adequate in isolation: we need to understand the dynamics of the interaction between the two levels.

The example shows that, while analysis at just one level, such as the social or the cognitive, can be in some circumstances be methodologically attractive, there can be important interactions between levels which should not be ignored. On the other hand, it is simpler to remain on one level, for example, to be concerned only with the social and not have to worry about the cognitive, or vice versa, and this is an adequate

methodological strategy if it does not do too much damage to the analysis. It would therefore be useful to have some rules of thumb about when one can analyse at just one level.

The most common reason for ignoring other levels is that the properties of the actors at these other levels can be assumed to be constant. When there is no endogenous change at the other levels and no 'leakage' from the level under analysis to other levels, it may be possible to confine one's attention to one level. Economists do this when they examine markets in equilibrium and assume that individual actors' preferences remain constant. They are duly criticised when the assumption is false, for example, when there are network effects such that the utility of a good changes because many other people have bought it. The famous example is the video cassette, where there was 'lock in' to the VHS format although the competing Betamax was technically superior, because there were overwhelming advantages to buying the same format as the majority of others had already chosen (Arthur, 1990).

A second reason for ignoring other levels is when there are many alternative processes at the lower level which could give rise to the same phenomenon at the level of primary interest. For example, a famous early social simulation was Schelling's model of residential segregation (Schelling, 1971). This model made only the crudest assumptions about the motivations of individual households and completely failed to examine why households might want to move out of neighbourhoods where they are in a minority. But for Schelling's purpose, which was to demonstrate the unintended consequences of the households' migration, the underlying motivations are not relevant and did not need to be modelled (Gilbert,

2002). This is fortunate because there are many reasons why a household might want to move. For example, members of a black household might want to move out of a white area because they wanted to be close to others of the same ethnicity, because they were priced out of an expensive white neighbourhood, because they were the victims of abuse and discrimination, or any number of other reasons. A multilevel analysis would need to examine and select between these motivations, but for Schelling, a study of them would have got in the way of the point he wanted to make.

Correspondingly, there are studies where it is impossible or unwise to confine the analysis to one level. Instances of these can be found throughout this volume. The most straightforward example is where the analysis of the lower level provides constraints on the phenomena at the higher level (see, for example, the chapters by Clancey et al and West, Lebiere and Bothell in this volume). There are restrictions on the speed of cognitive processing, on the physical location of bodies and on the physical possibilities of interaction, all of which impose constraints on the behaviour of social organisations, such as planning meetings and teams, but working out exactly what these constraints are and how they apply may require a very detailed analysis of cognition and even physiology.

A second case where one needs to model both the social and the cognitive levels is where the descriptions at one level are or can be applied at the lower level (what Sun calls 'mixed-level' analysis, see the chapter by Sun in this volume). An example often referred to in this volume is the Belief-Desire-Intention (BDI) cognitive model (Georgeff, Pell, Pollack, Tambe, & Wooldridge, 1998). This model seems to have been developed in the 1990s specifically to provide a plausible, yet computationally

tractable model of cognition for artificial intelligence research. It is not based on experimental evidence or on theoretical analyses of human cognition, but rather on what we might call ‘folk psychology’. Typically, in contemporary western culture we evaluate our peers’ actions by attempting to assess their intentions, referring to our knowledge or assumptions about their beliefs and desires. We do this without regard to either psychological knowledge about cognition, or, as philosophers (e.g. Winch, 1958) have pointed out, the logical puzzles that can arise when one tries to pin down ‘intention’ and separate it from ‘action’. Philosophical analysis has emphasised that the vocabulary of intentions is completely suffused with social action and cannot exist outside a social context. Thus, it is arguable that interpretations of action in terms of beliefs, desires and intentions are social constructs, products of a particular culture, and yet these have been pressed into service as a cognitive model by researchers. The BDI model is often useful, but perhaps no more so than a billiard ball model of the atom can be useful in understanding nuclear fission. Unlike some of the other cognitive architectures described in this volume, it is not corroborated by numerous experiments with human subjects under controlled conditions, but takes its plausibility from its effectiveness within our culture in making sense of others’ actions. In short, it is a model that mixes levels, using a socially constructed vocabulary to examine cognitive phenomena.

In this, the BDI model is a different kind of model from the others described in this volume, which are more firmly based on psychological theorising and experimental evidence. That still leaves the analyst with a choice to make, however. Should the social simulator opt for ACT-R or Soar, a connectionist neural network, CLARION, or some other model? Most of the contributors to this volume do not explain how

they selected which cognitive model to use, nor the implications of their choice. An exception is the chapter by West et al, who are explicit about the reasons why they believe ACT-R is the best model for their study: it is a good model of human cognition (but the proponents of other models would no doubt argue that theirs are good models also) and, more importantly, it was able to reproduce the pattern of sequential dependencies which they were aiming to model without ‘tweaking’ or special modifications. But as West et al note, the same is true of several other architectures. Here we have an example of where several cognitive architectures have the potential to generate an emergent phenomenon (the sequential dependencies) and so if one’s attention is focused on the social level, all the candidate architectures are more or less equally appropriate. In order to distinguish between them, one would need to identify cognitive or social level features where the models made different predictions and then see which of these is most in accord with data from observations of human individual and group behaviour. Until this is done, the choice of a cognitive model from among those described in previous chapters will probably depend on pragmatic issues such as how easy they are to obtain, previous experience with them and their use in related research.

I have argued that social simulations do not always need to be coupled to cognitive models. In some circumstances that I have begun to explicate, using cognitive models would only complicate the research. On the other hand, there are also studies where mixed level modelling seems inevitable given the approach taken, such as those that use BDI models, and other studies where a mixed-level analysis is essential. The benefits of a mixed level approach are apparent in several chapters in this volume, such as the one by Schur et al in this volume, where individual agents maintain

cognitive models (at the cognitive level) of the team as a whole (i.e. the social level), in order to improve the coordination of the agents' actions. Another type of example is provided in the chapter by Parisi and Nolfi in this volume in which they discuss very simple agents that exhibit behaviour similar to the flocking of birds or the schooling of fish, and show that this behaviour can be the result of the interaction of individual (which they call 'local') and social ('global') factors. Thus, the chapters in this volume provide some excellent examples of the problems and benefits of mixed level models, and we can look forward to more as the relationship between the social and cognitive levels is explored more deeply.

## References

- Arthur, W. B. (1990). Positive Feedbacks in the Economy. *Scientific American*, 262(February), 92-99.
- Baas, N. A., & Emmeche, C. (1997). *On Emergence and Explanation*. Santa Fe: Santa Fe Institute.
- Georgeff, M., Pell, B., Pollack, M., Tambe, M., & Wooldridge, M. (1998). The belief-desire-intention model of agency. In J. Muller, M. Singh & A. Rao (Eds.), *Intelligent Agents V*: Springer.
- Gilbert, N. (1995). Emergence in social simulation. In N. Gilbert & R. Conte (Eds.), *Artificial Societies: the computer simulation of social life* (pp. 144-156). London: UCL Press.
- Gilbert, N. (2002). *Varieties of emergence*. Paper presented at the Agent 2002 Conference: Social agents: ecology, exchange, and evolution, Chicago.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1, 143-186.

Winch, P. (1958). *The Idea of a Social Science*. London: Routledge & Kegan Paul.