

Co-creating Enduring Institutions for Socio-Technical Systems: The Complementarity of Content-based and Value-based Modelling Approaches

Simon T. Powers¹, Anikó Ekárt² and Peter R. Lewis²

¹Edinburgh Napier University, Edinburgh, UK, EH10 5DT

²Aston University, Birmingham, UK, B4 7ET

S.Powers@napier.ac.uk

As technology becomes ever more integrated into our daily lives, social issues become at least as important as technical ones. In other words, the systems are *socio-technical*. Many of these socio-technical systems provide new opportunities for citizens to work together to tackle pressing societal challenges. Examples include the use of artificial agents to automatically trade energy between users in smart grids, and the use of grid computing to harness the idle computing resources of millions of users to search for extra-terrestrial life.

But although these systems present great opportunities, they also present great challenges because they require individuals to cooperate by contributing their time, effort and resources to a shared enterprise. Consequently, they risk being subject to the Tragedy of the Commons where individuals act in a way that maximises their own payoff at the expense of the rest of the group.

Empirical work has demonstrated that some groups are able to avoid the Tragedy by co-creating and enforcing their own institutional rules that govern their interactions in the use of common-pool resources (Ostrom, 1990). But creating rules and then monitoring and enforcing their compliance is costly. So when will self-interested agents be incentivised to put the time and effort into creating and sustaining an institution? Ostrom (1990) derived a set of empirical principles for when this will occur. To implement these in systems that contain artificial as well as human agents, we need to formalise them and translate them into executable form. Previous work has done this using agent-based models (e.g. Pitt et al. 2012). Here we show how these can be complemented by an evolutionary game theory approach that can address questions of incentivisation in a rigorous way. Before doing so, we first draw out an important distinction between models of behaviour that are *content based* and those that are *value based*.

The Complementarity of Content-based and Value-based Models

Both agent-based modelling (ABM) and evolutionary game theory (EGT) are well-established approaches to modelling

social systems. We characterise these as instances of *content-based* and *value-based* modelling approaches, respectively. Figure 1 illustrates this.

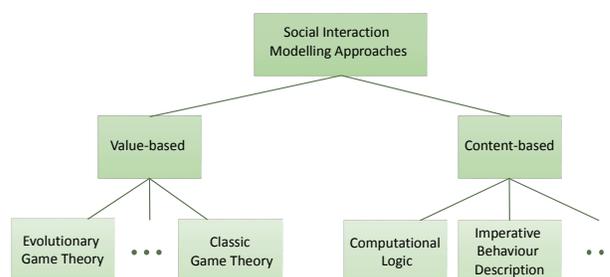


Figure 1: A taxonomy of approaches for modelling social interactions. The primary distinction is between approaches that capture the value of different strategies, and those that capture the behavioural content of those strategies. A range of game theory variants, including evolutionary game theory, can then be seen as value-based approaches. Agent-based modelling instead models the content of agent strategies. This can be implemented programatically in different ways as shown in the right-hand side of the figure.

In EGT the existence of a space of possible behaviours and their expected fitness is presented in a descriptive (equation-based) form. However, the *content* of the actions themselves that form part of the strategy, and lead to this fitness, are omitted. This omission includes any deliberative or developmental processes that are assumed to be included in the execution of the strategy; only the *value* of any such activity, in terms of fitness consequences, is given.

By contrast, in ABM we provide a description of the *content* of the modelled actions, typically in imperative or logical form, along with what effect they have on the world and other agents. Thus, it is possible to capture a deep and complex set of behaviours in an agent, based (for example) on learning, deliberative, and other cognitive processes. However, there is no explicit description of the *value* of carrying out the described activities, and furthermore, such a value is

hard to arrive at, save by executing the agent programs and observing.

In summary, each leaves implicit what is made explicit in the other. ABMs can capture rich behaviours, but struggle to support an analysis of their value. Conversely, EGT provides the necessary primitives to analyse the incentives and outcomes associated with different behaviours in a rigorous way, yet in doing so lacks the ability to capture what may be crucial details of the nature of the strategies themselves, and assumes that any value is accurately defined.

Predicting When Institutions Will Endure

We consider institutions to manage provision of a common-pool resource, similar to the setup considered in the content-based model of Pitt and Schaumeier (2012). Using an EGT model (Powers et al., in press), we focus on predicting conditions for the formation and maintenance of cooperation-promoting institutions, when individuals have to be incentivised to take on the institutional roles that are necessary for this, e.g. organising votes on rules or acting as a monitor. These predictions would be more difficult to make from a content-based model, since there is no direct currency of value in which to measure incentivisation. The calculation of critical thresholds and parameter values would require carrying out fully factorial parameter sweeps of executing the model, which is often not practical because of the amount of computation time required.

By contrast, using a value-based model we have derived analytical relationships between the model parameters that provide precise and easily interpretable answers to the following practical questions (Powers et al., in press):

1. How many individuals need to take on a monitoring role in order to incentivise cooperation?
2. How much should a group invest into monitoring in order to incentivise this number of individuals to become monitors?
3. What are the conditions for cooperation to become established given an initial state where no individual cooperates and no individual monitors?

Question 2 is particularly important, since knowing the answer avoids a group wasting resources by investing more into monitoring than is necessary. We provide the answer in terms of intuitive variables such as the number of individuals monitored by each monitor, and the time that it takes to monitor one individual.

The answer to question 3 is key when we are trying to help a group start managing its common-pool resources. Our results imply that some individuals will initially need to take on a monitoring role “for free”, since insufficient resources will be available to pay for them all. However, once cooperation becomes established then the group will be able to reimburse this charity by rewarding monitors with a greater share

of the common-pool resource. Our model shows this state is an equilibrium. But reaching this equilibrium requires individuals to be forward-looking to some degree, which is not captured in EGT models where individual cognition is assumed to be completely myopic. It could, though, be explored easily in an executable content-based model that implements cognitive theories of agent behaviour.

As such, the two modelling approaches can fill the gaps in each other. Future research should examine the extent to which EGT, or other value-based approaches, can be extended to capture more complex cognitive behaviour, where the value of a behaviour is not readily obtainable in general. One idea could be to induce the value of behaviours empirically, perhaps as a second layer in a content-based model. This might be done using evolutionary algorithms to give fitness values to evolving behaviours in a content-based model. This suggests that it might be easier to extract value from a content-based model than it is to add the content of behaviour to a value-based model.

We also need to determine how aspects of human psychology such as trust and fairness can be incorporated into models if we are to make better predictions. Is the best way to do this by assigning values to them, or do they defy value-based game theoretic modelling?

In the short term, however, we believe that it is important for modellers to provide clarity concerning whether their models either assume or explore the extent to which agents engage in cognition, or if they assume that agents simply ‘behave’. This is important, because model predictions may vary drastically as a result, and so this is necessary to provide the context for any resulting insight gained from the model.

Acknowledgements

This work was partially supported by a research grant from Edinburgh Napier University to S. T. Powers. We thank three anonymous reviewers for useful feedback and insight.

References

- Ostrom, E. (1990). *Governing the Commons : The Evolution of Institutions for Collective Action*. Cambridge University Press, Cambridge, UK.
- Pitt, J. and Schaumeier, J. (2012). Provision and appropriation of common-pool resources without full disclosure. In Rahwan, I., Wobcke, W., Sen, S., and Sugawara, T., editors, *PRIMA 2012: Principles and Practice of Multi-Agent Systems*, volume 7455 of *Lecture Notes in Computer Science*, pages 199–213. Springer Berlin Heidelberg.
- Pitt, J., Schaumeier, J., and Artikis, A. (2012). Axiomatization of socio-economic principles for self-organizing institutions: Concepts, experiments and challenges. *ACM Transactions On Autonomous And Adaptive Systems*, 7.
- Powers, S. T., Ekárt, A., and Lewis, P. R. (in press). Modelling enduring institutions: The complementarity of evolutionary and agent-based approaches. *Cognitive Systems Research*.