



# SCID

THE SOCIAL COMPLEXITY OF IMMIGRATION AND DIVERSITY

Social Complexity of Immigration and Diversity  
Review of Work

27<sup>th</sup> AND 28<sup>th</sup> OCTOBER 2011

**The Modelling Approach**

**by Bruce Edmonds**

**DRAFT PAPER – NOT FOR CIRCULATION OR CITATION**

## ***Modelling approach***

The SCID project aims to integrate social science understanding with complexity science tools and approaches in order to gain new insight into some of the social processes affected, directly or indirectly, by ethnic diversity and immigration. The interface between social science and complexity science is that of complex agent-based simulations. The idea is to build simulations that reflect some of the issues and evidence found in social science so that the complexity scientists can then analyse using their techniques. The approach chosen seeks to address some of the divisions and dilemmas involved in using simulation to represent aspects of society and hence develop a new way of understanding social phenomena.

The central dilemma in modelling social phenomena is rooted its sheer complexity – there is simply no guarantee that we will be able to understand any model that adequately captures what is being modelled. There are broadly two approaches to this difficulty: KISS (“Keep It Simple Stupid”) and KIDS (“Keep It Descriptive Stupid”) approaches (Edmonds and Moss 2005).

The former hope that progress will result from the formulation of relatively simple models that will approximate human social behaviour sufficiently well to be somewhat useful guides. Such models are sufficiently simple that their behaviour can be comprehensively analysed, possibly using analytic techniques. Such models are characterised by a use of random “proxies” for aspects of human social behaviour that are either poorly understood or too complicated. Another way of putting this is that the assumptions used are pretty strong and often controversial. These kinds of model are, in general, at some “distance” from observed social behaviour so that they tend to be used more as a model of a theory of behaviour, a way of explicitly representing and understanding a set of ideas about social mechanisms and how they might interact. What results is akin to an analogy, albeit in

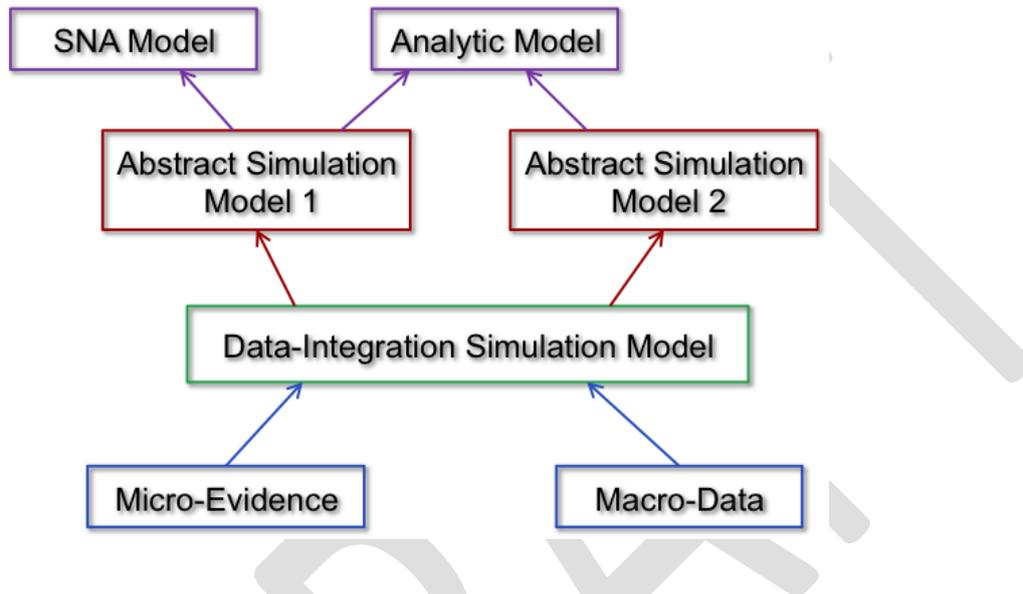
computational form, the model does not relate directly to the data but rather captures an idea which is then used to understand society in a rich, analogous but imprecise manner. To summarise, the KISS approach has rigour, in that the model can be well analysed and understood, but tends to lack close relevance to what is observed.

The second approach seeks to build simulations that more reflect what is observed to occur. Such an approach almost always results in relatively complex simulations, maybe with several social processes occurring and some representation of complex cognition in the computational representation of social actors. Here the assumptions tend to be of a weaker and more “mundane” type and the elements of the simulation correspond in a more natural way to what is observed. The simulations developed during the first stage of the SCID project (what will be discussed at the first annual review workshop) are of this type. The advantage of such simulations is that they can be related more directly to the available evidence at both the micro-level evidence as well as macro-level aggregate data – “cross-validated”(Moss and Edmonds 2005). The disadvantage of the KIDS approach is that the simulation models that result are slow and complex – sufficiently complex that one can never say that one completely understands the interactions within the model and almost always too complex to analyse analytically. Thus, although this approach produces an explicit model of the complex interaction of social processes, it does not result in a full understanding of them. To summarise, this approach has demonstrable relevance to what is observed but can be (rightly) criticised as lacking rigour, since one can never be completely sure that there is not a subtle error or inadequacy in the model.

The problem is that good science requires *both* relevance and rigour. It is fair to say that most social simulation modelling ducks this issue, but either accepting poorly understood simulations or else results that only weakly relate to the evidence. The SCID project aims at facing this problem head on. Its approach is not to use only one kind of model but to develop a “chain” of models of different

kinds, starting with complicated KIDS-style models and then progressing to KISS-style models.

Instead of attempting to represent observed social phenomena using KISS models, the KISS models will model the KIDS models which model what is observed. In this way we seek to stage the abstraction into “smaller” steps. The approach is illustrated in Figure 1.



**Figure 1. An Illustration of the SCID modelling approach, starting with the micro- and macro-level evidence and building “upwards” in stages to models that can be fully analysed.**

The advantages of this approach are that reference is explicitly maintained at each stage of the work: the complex “Data-Integration” model (DIM) is designed to capture what is observed in a relatively obvious and transparent manner, the DIM is then itself modelled using more abstract simulations whose behaviour and outcomes can be checked against experiments with the DIM, and finally even these fairly abstract simulations models might be approximated or further abstracted to network or analytic models checked against the abstract models. In a real sense what is observed is not modelled using a single model but by the whole chain. Every model can be checked against the lower levels from which it gains its meaning. The understanding of each level is confirmed by the analysis of levels above.

The disadvantages of the approach are twofold: it takes a lot of time and resources to develop such chains (hence the size and duration of SCID) and there is no well-established methodology of doing so. Developing, demonstrating, assessing and, if necessary, modifying this methodology is a main aim of the SCID project.

### ***Present Progress***

The first year of the project has been spent in developing the first DIM model. It was decided that this model would concentrate on the issue of why people vote, since this was an area of the greatest expertise among the project staff. Subsequent models on employment and trust issues are starting to be considered. It is anticipated that these will be developed far more rapidly due to the mutual familiarity and expertise gained through this year. Thus the substantive progress in this year has been the collation of evidence and formulation of the first version of the DIM model of the SCID project – the “voter model”.

To help focus the discussion about this model, we made a catalogue of underlying causal stories. Each of these was an explicit account of a process that might contribute (directly or indirectly) to the outcome of where someone voted. These are not a narrative<sup>1</sup>, in the sense that they have been collected from subjects, but rather they are a simple account from the experts involved in the project that accord with their opinion and the literature. They can not be called causes or causal theories since they would not qualify as a separable cause. Rather they are parts of the complex mix that results in the observed outcomes. Their import is that they help guide the programmer in terms of what processes are candidates for including in the DIM. Altogether we collected 54 of these stories, they are discussed below.

---

<sup>1</sup> We are looking to see if any of these are available, but have not found any yet.

One of the key advantages of DIM models is that they can include aspects of expert opinion and qualitative evidence as well as conclusions and data from quantitative studies. The qualitative evidence is most useful for informing the menu of possible social processes that occur at the micro-level, whilst quantitative data can be compared to the model during the validation phase. We have not yet done any independent validation of the model. Bridging the “gap” between rich qualitative narratives and the complex but formal micro-level behavioural rules of an agent in a simulation is an area of research that we intended to explore more rigourously during the SCID project. In lieu of this, the causal narratives are examples of what might result from an analysis of narrative accounts from stakeholders. However, these narratives do have evidence to support them, and have been found to be acceptable to a range of domain experts, so they provide a good starting point. The explicit declaration of such causal stories is important in order to make the model as criticisable as possible, and open to future amendment and improvement.

The modelling progressed with frequent meetings between the modelling (CPM) and domain experts (ISC), up to twice a week. Thus the model developed with a continually high level of input from the evidence as presented with and discussed by the experts. Although, this did not go as far as participatory approaches to model building, this is in line with making the DIM as evidence-led as possible, rooted more in the available social science rather than as a guess/construct by a computer scientist. The point is not to obtain to a perfect model, but rather to explore a rich mixture of social processes which are explicitly documented. This will allow future versions of the model to be developed and compared with the original.

The next stages of the SCID project will include: validating the DIM against as wide a set of data sources as possible, exploring variations of the reference DIM, building more abstract models that align with the DIM, and finally using the whole chain of models to shed light on policy issues.

### ***Reference list***

Edmonds, B. and S. Moss (2005). "From KISS to KIDS—an 'anti-simplistic' modelling approach." Multi-Agent and Multi-Agent-Based Simulation: 130-144.

Moss, S. and B. Edmonds (2005). "Sociology and simulation: Statistical and qualitative cross-validation." American Journal of Sociology **110**(4): 1095-1131.

DRAFT