Using a Data-Integration Model to Stage Abstraction in Modelling Occupational Attainment

Bruce Edmonds¹, Ruth Meyer¹ Huw Vasey²

1 Summary
The “Social Complexity of Immigration and Diversity” (SCID) project applies agent-based simulation techniques to understand the impact of immigration and diversity on social integration, cohesion and inequality. In particular, a chain of simulations – from complicated and descriptive up to abstract and tractable – will be used to bridge the explanatory gap between micro- and macro-level evidence, allowing for a more complete and coherent understanding of the social complexity associated with immigration, social integration and diversity. In this paper, we present our approach to modelling and outline the first stage of abstraction in the labour market case study, a data-integration model of occupational attainment.

2 Introduction
Although mindful of the economic and cultural benefits of immigration, governments are concerned about problems of the social, economic and political integration of immigrants, and native citizens’ responses to an increasingly diverse society (CANTLE, 2001; DENHAM, 2001). Substantial evidence exists in many western democracies of immigrant disadvantage in labour market opportunity, social mobility, educational achievement and political participation (LI & HEATH, 2007). Theoretical accounts for this range from assimilation to multicultural pluralism. Some (e.g. ALBA & NEE, 2003; PARK & BURGESS, 1921) are essentially optimistic, suggesting that disadvantages will disappear with acculturation. Other theories (e.g. GORDON, 1964) argue that even if acculturation comes about disadvantage will persist until entry into the “cliques, clubs, and institutions of society” has occurred (GORDON, 1964, p.71).

When beginning to analyse the social complexity of immigration and diversity, the structures of the labour market (social, legal, geographic and economic) immediately appear to play a central role. What jobs people get – immigrants and ethnic minorities included – can be seen as a central driver of both immigration and diversity. That is, issues such as whom you work with, how much you earn, where you work and so forth, have a fundamental impact on other areas which affect social diversity, such as where you live and what your social networks are.

However, the processes by which this works are often hidden. Whilst excellent studies have illustrated the presence of ethnic segmentation in the labour market and its relationships with both international migration and social and economic diversity (BAUDER, 2006; McDOWELL et al., 2007, 2009; WALDINGER & LICHTER, 2003, WILLS et al., 2010), how this occurs is unclear. We then rely on socio-

¹ Bruce EDMONDS, (email: bruce@edmonds.name), Ruth MEYER (email: r.meyer@mmu.ac.uk), Centre for Policy Modelling, Manchester Metropolitan University, Oxford Road, The Business School, Manchester, M15 6BH, UK
² Huw VASEY (email: huw.vasey@manchester.ac.uk), Institute for Social Change, Manchester University, 2.11 Humanities, Bridgeford Street, Oxford Road, Manchester, M13 9PL, UK
logical theories to make assumptions which, while considered and internally coherent, are often difficult to test empirically.

For example, there is a large volume of substantive evidence regarding the persistence of ethnic segmentation in the labour market beyond first generation immigrants (Heath, 2007). However, studies inherently find it difficult to investigate the (hidden) mechanics of the interactions between discriminatory behaviour by employers (conscious or otherwise), legal rules and processes (such as employment law and immigration rules), and the dual interpellation of affect and ambition between employer and employed, society and individual.

Developing a detailed, descriptive agent-based model provides a chance to do this.

The Modelling Approach

The SCID project (SCID, 2010) 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 in 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 & 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 & 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 rele-
vance to what is observed but can be (rightly) criticised as lacking rigour, since one can never be com-
pletely 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, by either accepting poorly understood simulations or 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.

![SCID Modelling Approach Diagram]

**Fig. 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.

### 3 The Employment Model

The first version of a DIM describing labour market interactions in the UK is currently being devel-
oped. It is intended to investigate the processes of occupational attainment across different ethnicities.
That is, what jobs people get and how and why they get them.

Recent social research has indicated a) that occupational attainment is different (even when factors,
such as level of education, are taken into account) along certain lines – such as parental occupation
(class), ethnicity and gender, and b) that there are strong social processes – networks of friends, acquaintances and employers; conscious and unconscious assumptions about the ‘nature’ of others and ourselves; legal regimes and restrictions –, which have a profound impact on our abilities to get certain jobs in certain places.

This model is an attempt to produce a simulation of the multi-faceted interactions which make up the labour market; creating rules which mimic each micro-interaction based on available substantive data. These rules, in turn, are based on an analysis of the key domains of influence which affect our occupational attainment throughout the life course (see Figure 2).

![Fig. 2: The key domains influencing occupational attainment and their interrelations. Macro-level domains are coloured grey, micro-level domains are white.](image)

The DIM represents both individuals seeking labour (employees) and companies offering jobs (employers) as agents, whose behaviour is governed by sets of rules derived from empirical evidence. On the macro-level, laws and regulations concerning e.g. discrimination and ethnic equality may constrain the actions of agents on the micro level, e.g. influencing the staff selection strategies of employers. The interaction of agents through their behavioural rules will then produce a model of the labour market, which can be adapted to a variety of scenarios – for example, by changing the rate of international migration into the model – which will allow us to ‘test’ how certain changes affect the model.

Using data sources including Understanding Society (the UK Household Longitudinal Study), Labour Force Survey, ONS Vacancy Survey and ONS VAT Registry we were able to build a reasonable approximation of the UK job market in 2012. Initialising the model with ten thousand agents of working age, the modelled job market consists of 8239 job “slots” of six different types (managerial/professional, intermediate, routine in both full-time and part-time) in 665 organisations, ranging in size from sole traders to 500+ employees. Four industry sectors are differentiated: construction, public sector and health, production and manufacture, and services, with the last one being the largest and most diverse – offering jobs from part-time routine roles in small shops, to full-time managerial/professional roles in large employers.
4 Conclusion and Outlook

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 rigorously during the SCID project.

The modelling is progressing with frequent meetings between the modelling team and domain experts, up to twice a week. Thus the model is developed with a continually high level of input from the evidence as presented with and discussed by the experts. Although this does not go as far as participatory approaches to model building, it is in line with making the DIM as evidence-led as possible, rooted more in the available social science rather than plausible assumptions by a computer scientist. The point is not to obtain 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.

5 References


