MANCHESTER 1824
The University of Manchester

Manchester Metropolitan University

SCID

THE SOCIAL COMPLEXITY OF IMMIGRATION AND DIVERSITY

The General Modelling Approach

Institute for Social Change &
Theoretical Physics Group,
University of Manchester

Centre for Policy Modelling,
Manchester Metropolitan University

EPSRC
Engineering and Physical Sciences Research Council

Outline of Talk

1. The Micro-Macro Link as motivation
2. A bit about modelling in general
3. A bit about “Complexity Science”
4. The central dilemma: *simplicity or not*
5. The proposed way forward: “chains” of increasingly abstract/simple models
6. About Data-Integration Models
7. The stage we have got to in SCID

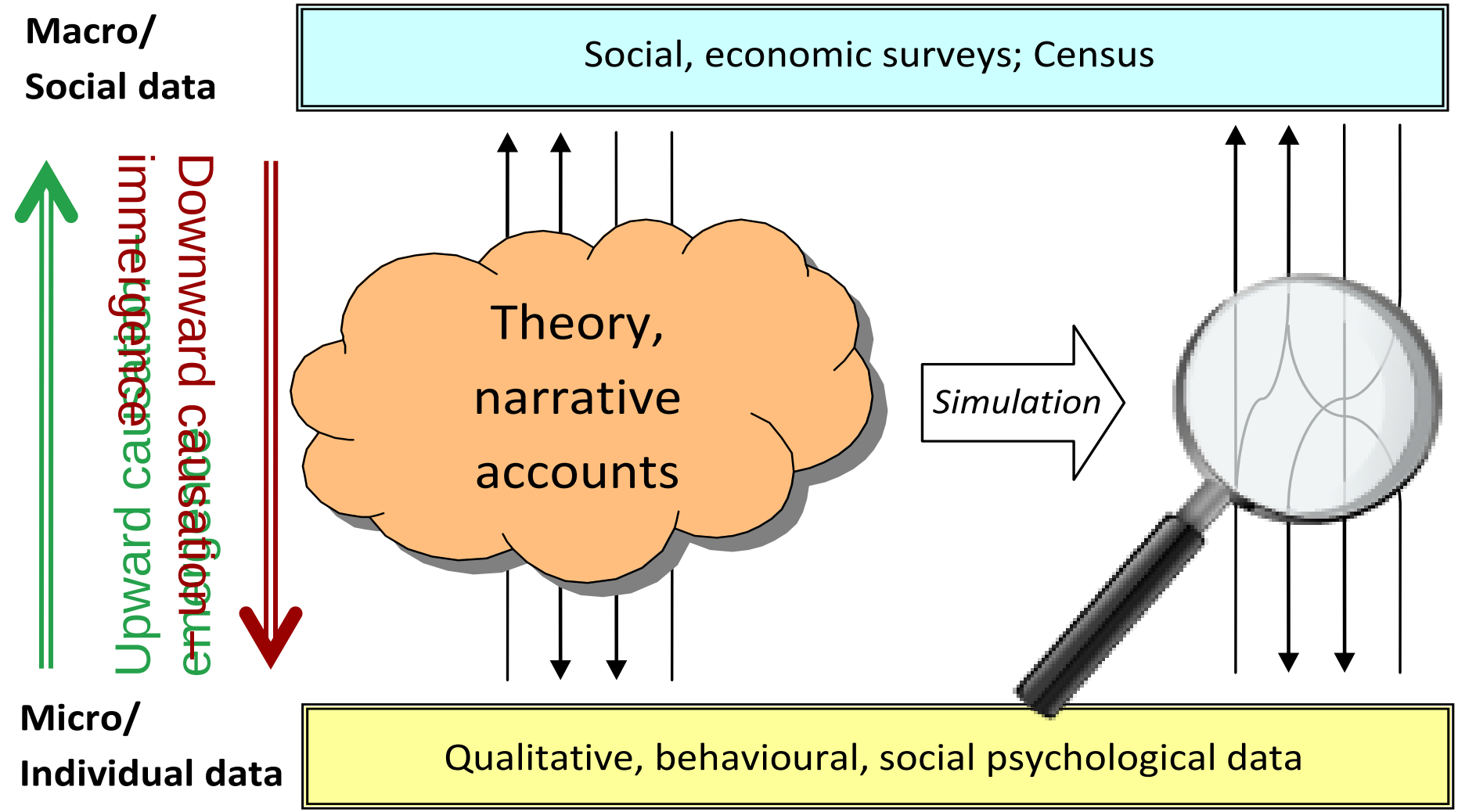
The Micro-Macro Link

- How do the tendencies, abilities and observed behaviour of individuals…
- …relate to the measured aggregate properties of society?
- Social Embedding etc. implies this link is complex
- Averaging assumptions (a general tendency + random noise) do not capture non-linear interaction
- This is often two-way, with society constraining and framing individual action as well as individual constituting society in an emergent fashion
- Somewhat-persistent, complicated meso-level structures mediate these effects – these might be key to understanding this

In Vitro vs In Vivo

- In biology there is a well established distinction between what happens in the test tube (*in vitro*) and what happens in the cell (*in vivo*)
- *In vitro* is an artificially constrained situation where some of the complex interactions can be worked out…
- ..but that does not mean that what *happens in vitro* will occur *in vivo*, since processes not present *in vitro* can overwhelm or simply change those worked out *in vitro*
- One can (weakly) detect clues to what factors might be influencing others *in vivo* but the processes are too complex to be distinguished without *in vitro* experiments

Unravelling the Micro-Macro Link



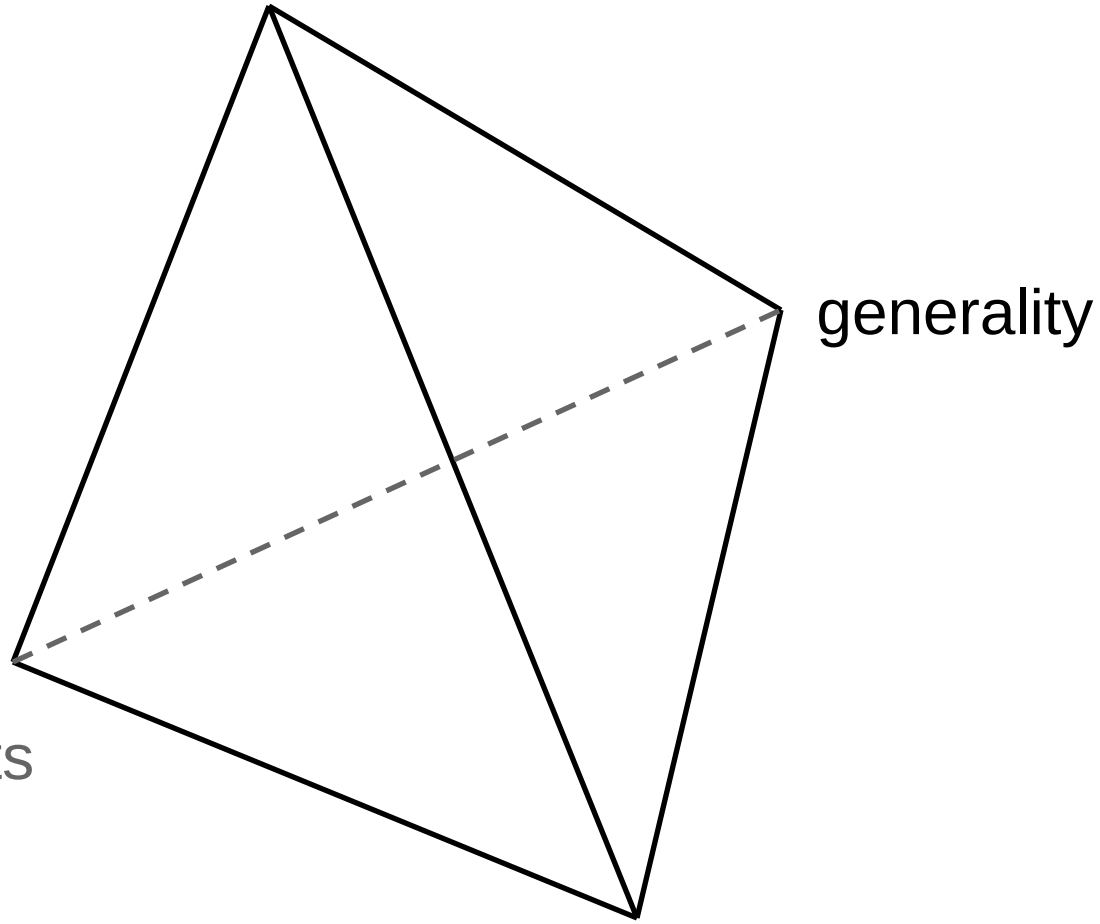
Kinds of Model

- Discursive – expressed in natural language using analogy, narratives etc.
- Data sets – sets of abstracted observations usually numerical
- Analytic mathematics – sets of equations that can be solved for a general solution
- Statistical models – estimations and tests based compared with the random case
- Simulations – the model is a computer program whose outcomes are computed

Some criteria for judging models

- Soundness of design
  - w.r.t. knowledge of how the object works
  - w.r.t. tradition in a field
- Accuracy (lack of error)
- Simplicity (ease in communication, construction, comprehension etc.)
- Formality (lack of ambiguity)
- Sensitivity (relates to goals and object)
- Plausibility (of design, process and results)
- Cost (time, space etc.)

Some modelling trade-offs



"Complexity Science"

- There is no single "Science of Complexity" but…
- ..rather this is an umbrella term covering several approaches and fields…
- …where the complicated interaction of many parts can produce "surprising" outcomes
- In particular, where neat analytic models and solutions are not feasible…
- …hence the preponderance of "individual-based simulations" (each part kept track of separately)…
- …which are "agent-based simulations" in the social sciences since the processes in the individuals are interpretable as some kind of cognition

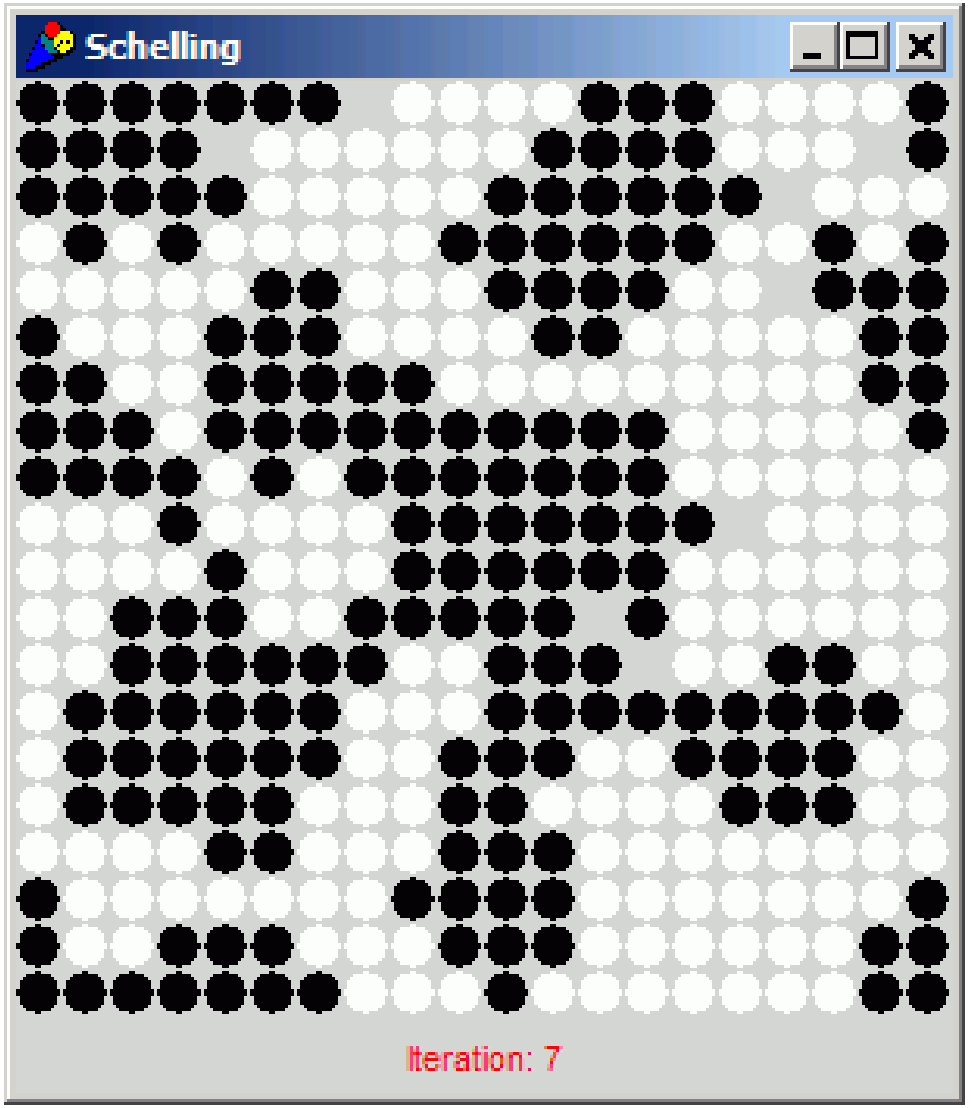
A Classic Example: *Schelling's Segregation Model*

Schelling, Thomas C. 1971. Dynamic Models of Segregation. *Journal of Mathematical Sociology* **1**:143-186.

Rule: each iteration, each dot looks at its neighbours and if less than 30% are the same colour as itself, it moves to a random empty square

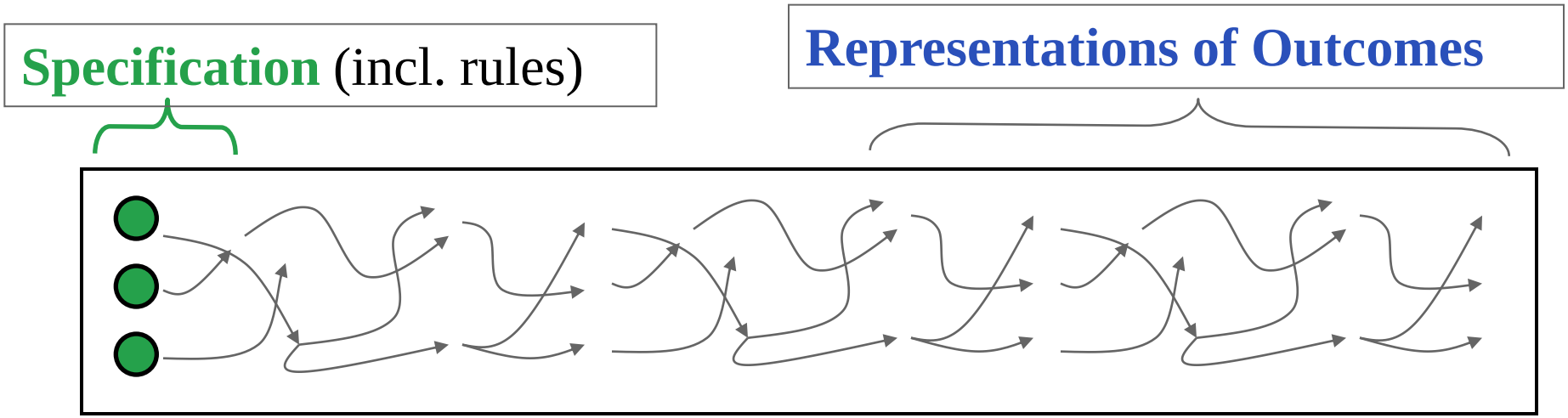
Conclusion:

Segregation can result from wanting only a few neighbours of a like colour



What happens in an Agent-Based Social Simulation

- Entities in simulation are decided
- Behavioural Rules for each agent specified (e.g. sets of rules like: if *this has happened* then *do this*)
- Repeatedly evaluated in parallel to see what happens
- Outcomes are inspected, graphed, pictured, measured and interpreted in different ways



Complexity in Practice

- Has classically focused on how the interaction of fairly simple parts can result in complex structures/outcomes: i.e. the phenomenon of *emergence*
- The “embedding” of the parts in their local situation is crucial to resulting outcomes and explanations
- Has resulted in lots of new approaches & methods for analysing data & models
- Hence the physicists in SCID, however some traditional habits persist, e.g.
  - Desire for simple and general models
  - Focus on complete understanding of single models
  - Need for analytic results to be publishable

Interfacing Physics and Social Science Approaches

- Physics and Social Science have very different languages, cultures and approaches
- We would like the power of approaches and tools of complexity physics but appropriately applied and not in “brave leaps” of abstraction which lose relevance to the observed
- (In particular the way that much work in economics involves unrealistic assumptions and a lack of relevance to what is observed)
- Thus in SCID simulations, albeit complex ones, will be the common interface and provided a common reference

The Central Dilemma

- *The social world we study is not arranged for our convenience (as academics)*
- Corollary… there is no reason to suppose that: *a model adequate to its representation will be simple enough for us to understand*
- There are reasons to suppose that it is not (*social embeddedness, SIH, Machiavellian Intelligence, cognitive competition etc.*)

KISS vs. KIDS

- **Keep It Simple Stupid**: Models that are simple enough to understand and check (rigour) are difficult to directly relate to both macro data and micro evidence (*lack of* relevance)
- **Keep It Descriptive Stupid**: Models that capture the critical aspects of social interaction (relevance) will be too complex and slow to understand and thoroughly check (*lack of* rigour)
- But we need ***both*** *rigour* and *relevance*
- Mature science connects empirical fit and explanation from micro-level (explanatory and phenomenological models)

The Resulting Dilemma

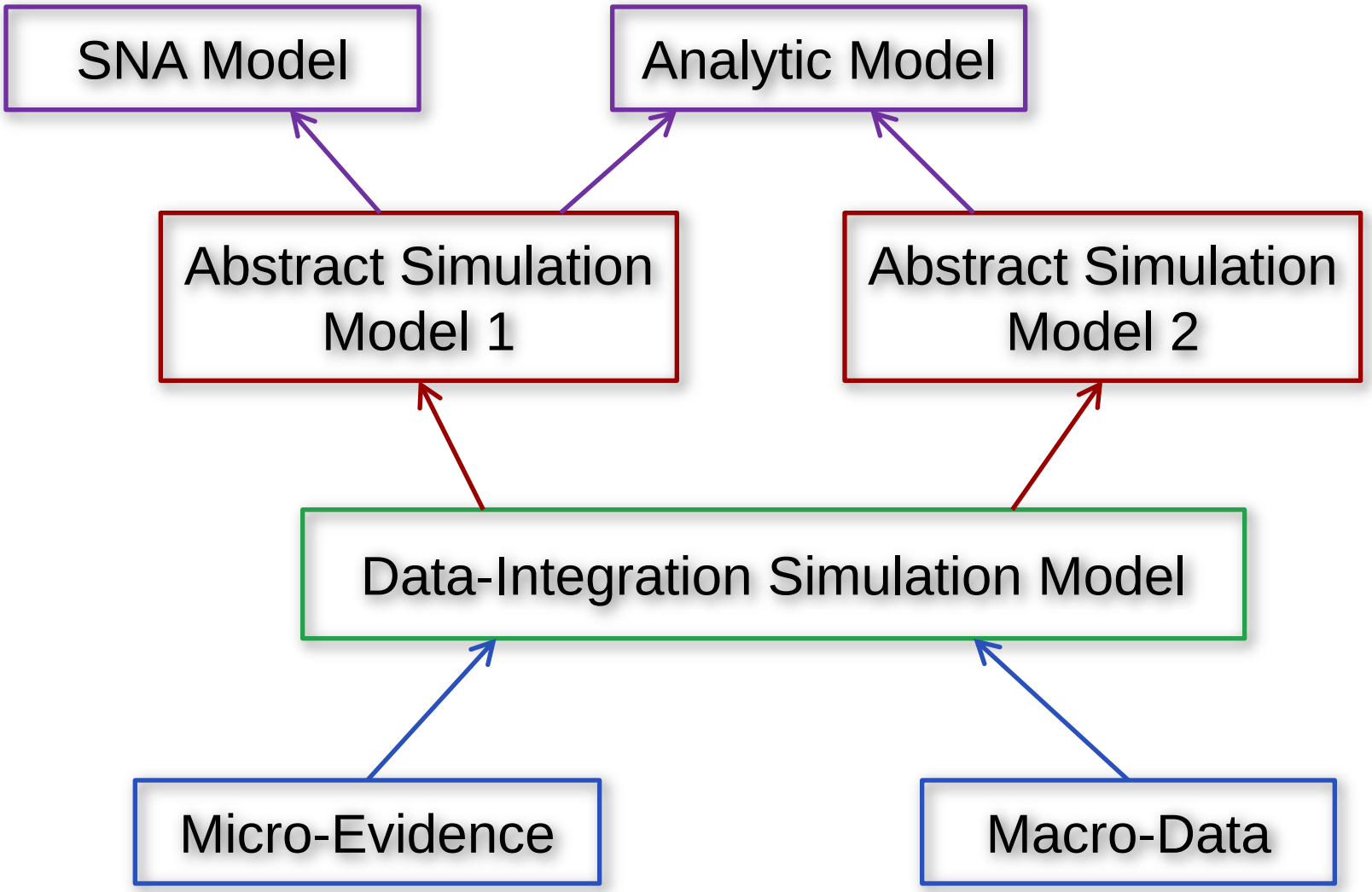
- **KISS**: Models that are simple enough to understand and check (rigour) are difficult to directly relate to both macro data and micro evidence (*lack of* relevance)
- **KIDS**: Models that capture the critical aspects of social interaction (relevance) will be too complex and slow to understand and thoroughly check (*lack of* rigour)

- But we need ***both*** *rigour* and *relevance*
- Mature science connects empirical fit and explanation between micro- and macro-levels (explanatory and phenomenological models)

The Proposed Approach

- Not to use a *single* model but rather a closely related *"chain" of models*
- Starting with narrative and statistical evidence for the micro-level behaviour of individuals etc.
- To build models that are more adequate to the processes that are thought to occur
- Which are checked and assessed against as many kinds of evidence as possible (including macro statistical evidence)
- And only later abstract to simpler simulation, analytic and social network models

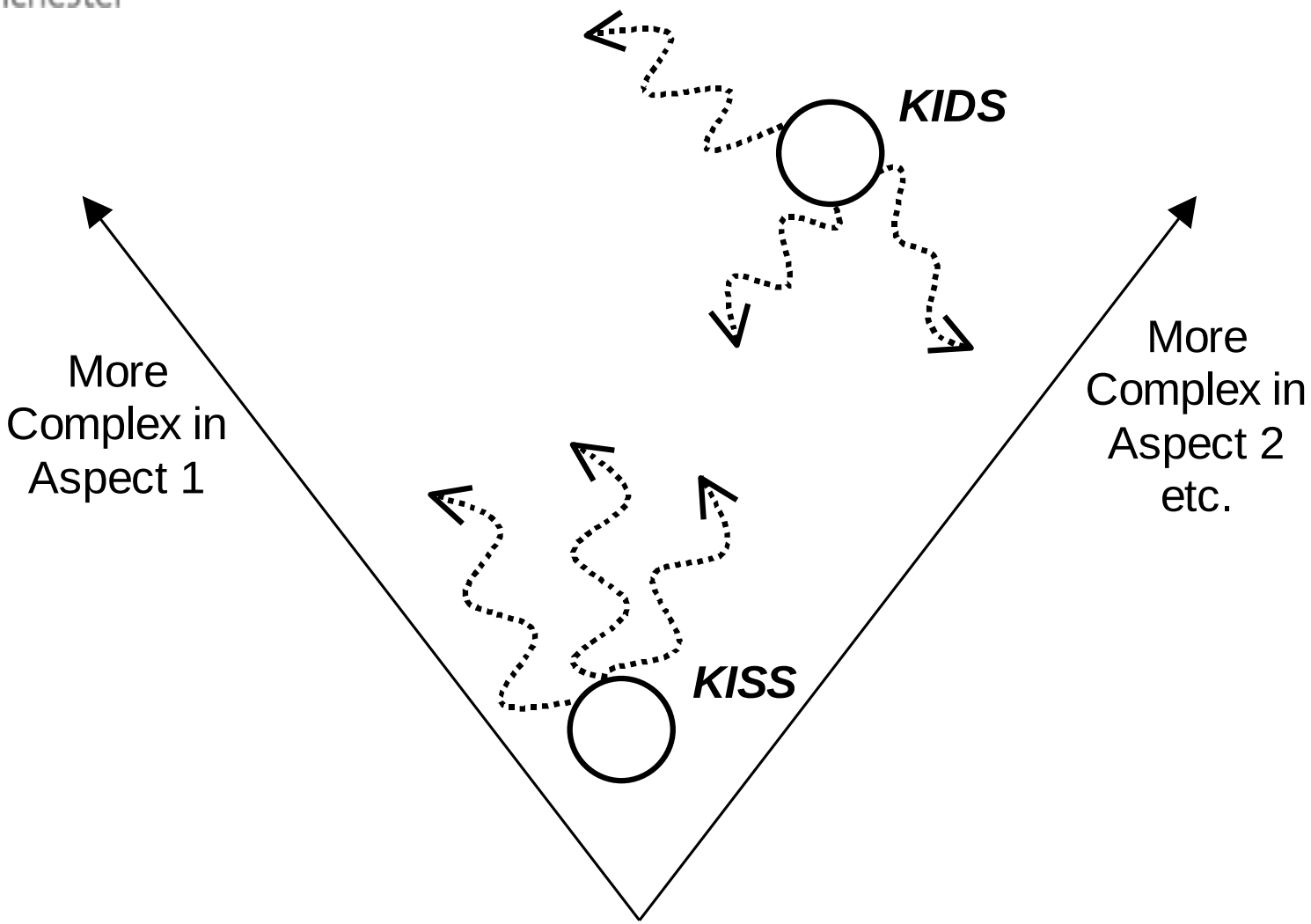
the SCID Modelling Approach



Roles of each kind of model

- Each is constrained by those “beneath” them, i.e. are consistent with them
- What each component should clearly represent something
- Models “above” analyse, check and explain what is happening in those below
- Models immediately “below” can be used to explore the safety of assumptions
- It might well happen that simpler, more abstract models have validity (w.r.t. a lower model) only under some settings

KISS vs. KIDS as a search strategy



But why not just jump straight to simple models?

- There are many possible models and you don't know *why* to choose one rather than another, this method provides the underlying reasons
- Much social behaviour is context-specific, and this approach allows one to check whether a particular simple model holds when background features/assumptions change
- The chain of reference to the evidence is explicit, allowing one to trace their effect and possibly better criticise/improve the model
- This approach facilitates the mapping onto qualitative stories/evidence

Data Integration Models

- Intended more as a computational description of a particular case than a theory (at least a general theory)
- Its aim is to represent as much of the relevant evidence as possible in one coherent and dynamic simulation
- Provides a precise target for abstraction (which are then checkable against it)
- *Stages* abstraction from data to theory
- Separates representation and abstraction
- Preserves chains of reference

Aims and Objectives of DIM

- To develop a simulation that integrates as much as possible of the relevant available evidence, both qualitative and statistical (a Data-Integration Model – a DIM)
- Regardless of how complex this makes it
- A *description* of a specified kind of situation (not a general theory) that *represents* the evidence in a *single, consistent and dynamic simulation*
- This simulation is then a fixed and formal target for later analysis and abstraction

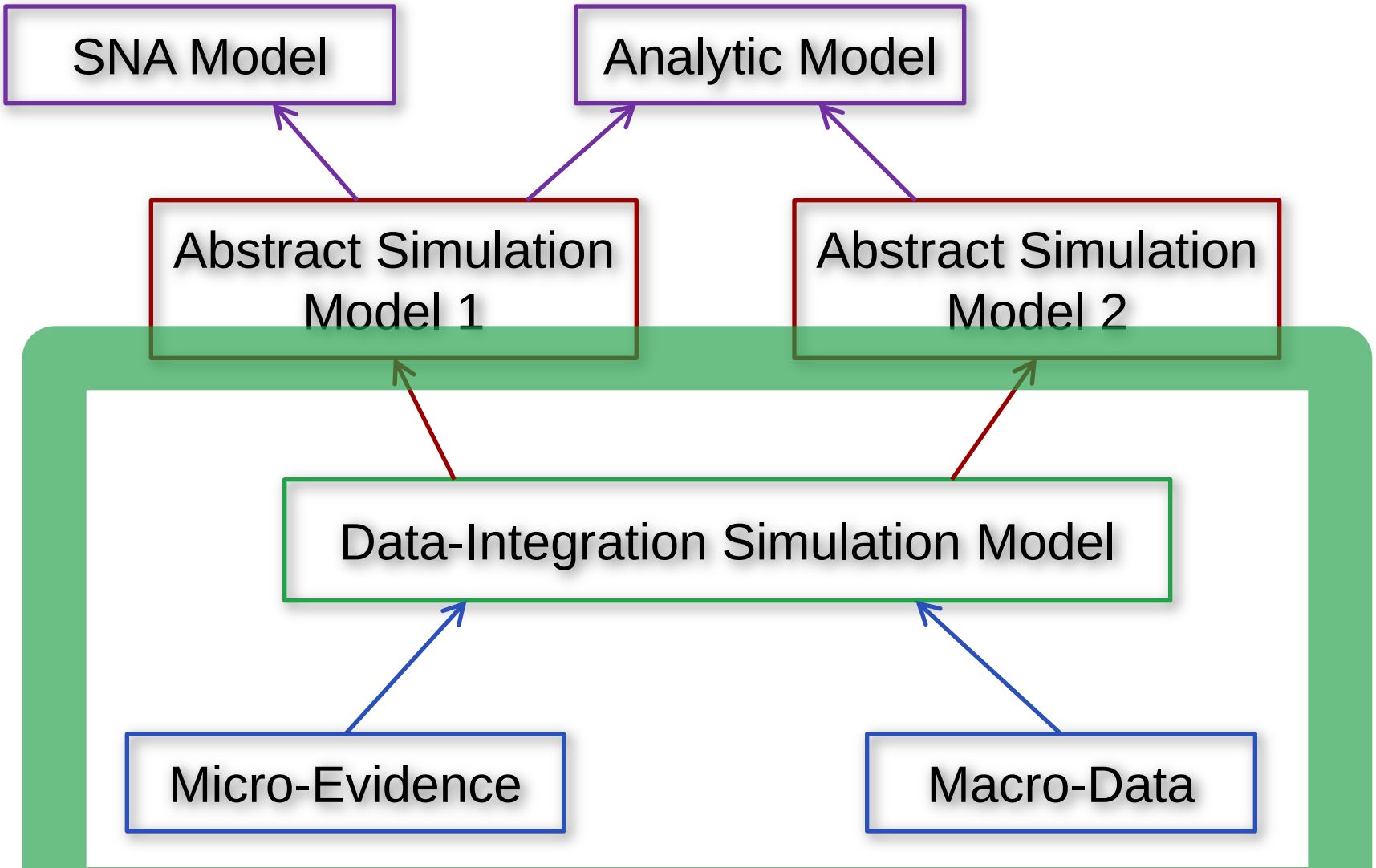
DIM Development Method

- A relatively tight interactive “loop” between the social scientists who are experts in the subject matter and their data and the simulation developers...
- ...trying to give as much ownership and control to social scientists as possible.
- First target: *What makes people vote (within the context of a diverse community)?*
- Started with developing a fairly complete list of “causal stories” concerning the various processes that might contribute from
- Then initial model iteratively developed in NetLogo to enable maximum responsiveness and transparency
- To be reimplemented in Java/Repast when the target becomes more “settled” for more extensive simulation exploration and analysis

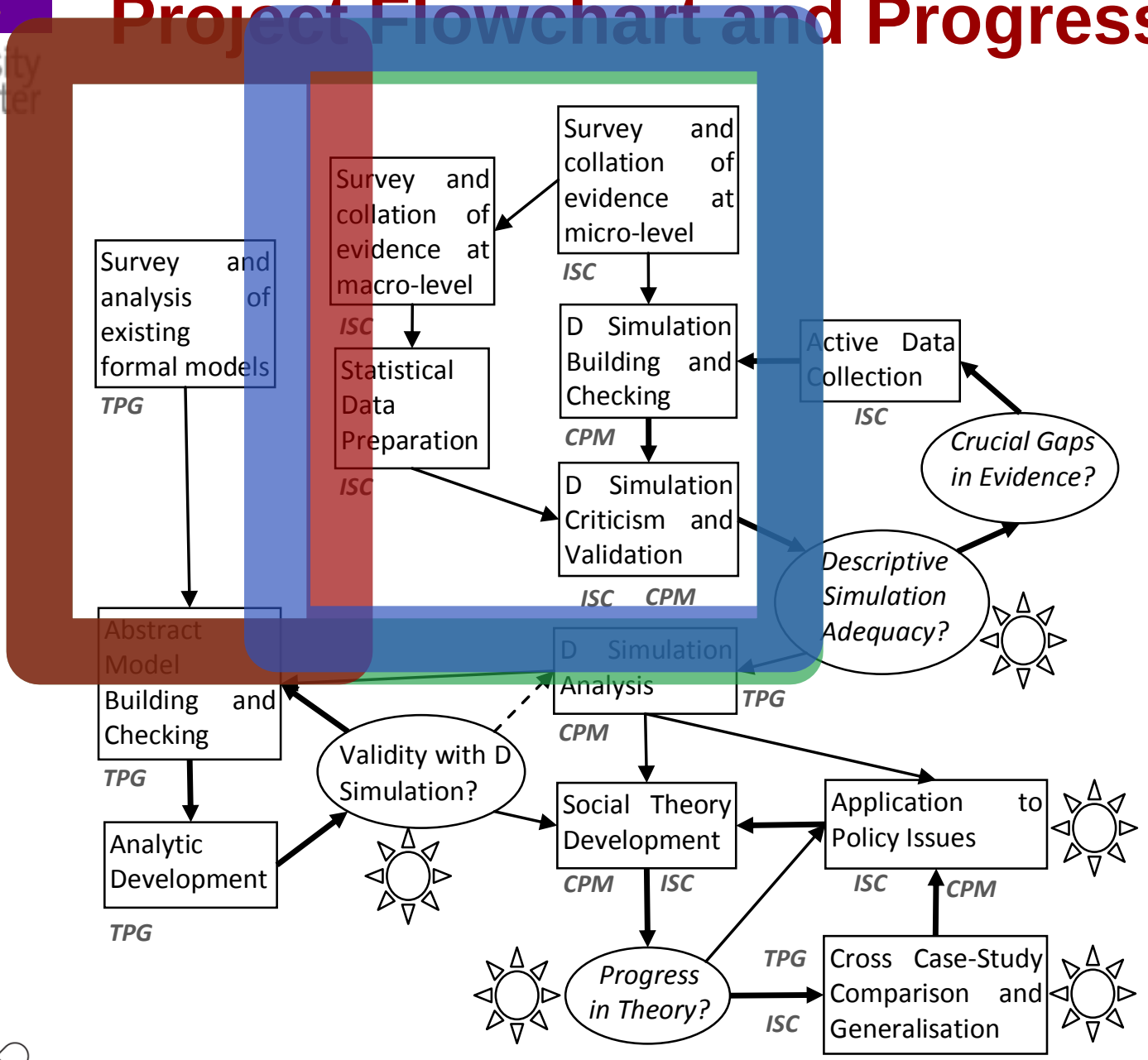
Progress So Far

- First DIM model produced on voter behaviour within a social/demographic context
- Validation, adjustment and exploration of DIM yet to occur
- Gaps in knowledge and hence future data collection prepared for
- Analysis, simplification of DIM by Theoretical Physics team yet to occur
- Policy analysis still a long way away

Recap of the Modelling Approach



Project Flowchart and Progress



Methodological Issues Arising

- Development of *qual→agent rule* process needs more research so it is more systematic, replicable and transparent
- Maybe better worked out criteria of what to include in a DIM
- Shows need for a library of ‘mundane’, underlying agent-based models of, say, population, household structure change
- Descriptive/diagramming techniques to make simulation design more accessible to non-programmers

The End