Why Sociology Should Use Agent Based Modelling

by Edmund Chattoe-Brown
University of Leicester

Sociological Research Online, 18 (3) 3
<<http://www.socresonline.org.uk/18/3/3.html>>
10.5153/sro.3055

Received: 13 Feb 2013 Accepted: 23 Apr 2013 Published: 31 Aug 2013

Abstract

Although Agent Based Models (hereafter ABM) are now regularly reported in sociology journals, explaining the approach, describing models and reporting results leaves little opportunity to examine wider implications of ABM for sociological practice. This article uses an established ABM (the Schelling model) for this. The first part argues that ABM integrates qualitative and quantitative data distinctively, provides novel tools for understanding social causes and offers a significantly different perspective on theory building. The second part shows how the emerging ABM methodology is compatible with existing sociological practice while undermining several criticisms of ABM perceived to limit its sociological relevance.

Keywords: *Agent Based Models, Complexity, Residential Segregation, Schelling Model, Structuration, Theory Building*

Introduction

1.1 Rather than presenting results from a novel ABM (which now occurs more frequently in sociology^[1]), this article describes a pre-existing ABM but analyses its ramifications for sociological practice. It begins by illustrating three ABM advantages (novel combining of qualitative and quantitative data, transcending linear notions of cause and effect and building distinctive theories). The article then shows how emerging ABM methodology (which is less well understood in sociology) augments existing sociological debates and also demonstrates that some criticisms of ABM don't impair its sociological usefulness.

The Schelling model

2.1 This example (Schelling 1969) has been deliberately chosen for simplicity. Since the article's point is to analyse the ramifications of ABM for sociology (which can be done equally well with a simple model), a novel ABM was purposely not presented in case the reader's attention was diverted from the implications of the *approach* to the implications of the model. However, as the article will show, features of the Schelling model which sociologists may find unappealing (lack of empirical grounding, simplistic and arbitrary assumptions) *aren't* intrinsic to ABM. It is very important for the reader to understand that I don't 'believe' the Schelling model assumptions or 'endorse' its outcomes. It is *only* being presented as an easily described example to underpin my argument.

2.2 Imagine a world comprising a regular grid like graph paper. Each grid site can be in one of three different states. It can be empty or occupied by a single 'social actor' (called 'an agent' for reasons to be explained) of one of two different types. Schelling was interested in residential ethnic segregation but, avoiding implications of realism, these types will be called squares and triangles. To initialise the simulation, agents (of both types) and empty sites are randomly distributed on the grid.

2.3 Schelling made stylised assumptions about how agents act. (Keep clearly in mind that these *are* stylised assumptions.) Each agent only 'regards' the state of eight directly neighbouring sites in deciding how it 'feels' about its location. It is further assumed that empty sites have no effect on this 'feeling'. Thus, in this model, the only thing that matters to an agent about its location is what the neighbouring types are. Possibilities thus range from being entirely surrounded by own type agents to being entirely surrounded by other type agents. An agent attribute called Preferred Proportion (hereafter PP) is defined

as the *minimum* fraction of one's own type that must be neighbours for an agent to be 'contented'. To complete the assumptions it is necessary to explain why 'contentment' matters. A 'content' agent will stay still. An agent that is not will move to a location chosen randomly from all vacant sites ('hoping' to find 'contentment').

2.4 What should the reader make of this? Clearly, it is significantly different from qualitative and quantitative articles chiefly comprising sociology journals. Recall that this model has *not* been presented for assumption plausibility or result relevance. For now it is pure formalism without claims on reality. Certainly terms like 'contented' and 'feel' have been used but in scare quotes and to facilitate understanding. I have no interest in sliding claims that the Schelling model is plausible past the reader. The article simply states what it assumes. ABM plausibility will be carefully considered later.

2.5 For now, the important thing is the computer's role and thus what is distinctive about ABM. The article refers to 'agents' because how the computer programme is written distinguishes it from other (typically older) simulation approaches. ABM instantiate a key sociological intuition, that agency resides in entities like humans but not in rocks. Implementing Schelling's model as an ABM involves agency residing where it 'should' (namely in programme elements specified discretely as agents which 'do things' like decide and move). In other simulation approaches (for example System Dynamics – Forrester 1971) the objects that the computer operates on only represent the social world *indirectly* (via social science entities like transition probabilities – Bartholomew 1973 – or differential equations). In such approaches agency and causality become unclear. When an equation states that unemployment each year is 5% more than it was last year, what is it saying about how unemployment arises, why it is increasing and who is implicated? Of course, representing social world ontology 'directly' does not guarantee sociological understanding but, in doing this, ABM need not immediately repulse sociologists approving of rigour but finding inadequate evidence that equations can represent human behaviour.

2.6 Presenting matters thus, the computer's role in ABM is less alienating than sometimes supposed. Like statistical software, it is possible to know in some detail what is happening without knowing exactly how the computer achieves it. Specifying 'rules' for Schelling's model can be done *without* a computer. (A short narrative has been used here.) The computer's role is to implement the rules as a programme and run it fast and reliably so that the resulting dynamic process can actually be observed. The considerable significance of what may seem a trivial contribution will be explained below.

Implication 1: Another 'lens' on society

3.1 It is impossible to avoid having preconceptions about society. These may originate in (or perhaps make us favour) particular research methods. One view (matching quantitative research) is that a smallish number of social categories (like class and gender) identify social differences (like voting or educational attainment). Another view (matching qualitative research) sees subjective accounts and environmental detail as crucial to understanding social life. In the latter view, detail isn't 'random noise' obscuring statistical relationships between variables but the very substance of sociology. From a qualitative perspective, quantitative variables need replacing by observation of social interactions in which people 'do' gender (or whatever) if *explanation* is to be achieved (about *how* girls and boys succeed in school for example). The rest of this article will show that these are not (as often implied) methodological positions whose assumptions cannot be evaluated empirically. In fact, ABM suggests how they might simultaneously be combined and tested.

3.2 The implications of these views (stereotyped for concision though they are) are recognisable in sociological practice. The quantitative inclination (also seen in approaches like rational choice) is simplification: producing hypotheses suiting available data.^[2] The qualitative inclination is discovering detail, celebrating diversity and critically deconstructing quantitative 'simplifications'.^[3]

3.3 However, each view contains recognised sources of weakness. The potential cost of simplification is irrelevance. A model may fit the data (or be analytically tractable) but cease to apply to (or reveal anything about) society.^[4] Conversely, if too many social categories are relevant to a given social difference, quantitative analysis needs too much data to reach trustworthy conclusions. Dissecting social outcomes via quantitative categories may produce not insight but muesli, depending on how complicated the world actually is and in what ways. Conversely, qualitative analysis may become so focused on describing detail that it neglects evidential criteria for generalisation, producing subjective narratives with the same scientific status as journalism.

3.4 What can be done about this? One answer is that novel research methods might help with weaknesses of existing methods (though clearly they will bring their own weaknesses.) The article will now show (using Schelling's model) that ABM is such a novel method, framing sociological problems and research processes significantly differently.

3.5 The first illustration involves considering what is needed to convert Schelling's model from a 'toy' (pursuing the implications of arbitrary assumptions without calibration and validation) to a 'proper' ABM. (Proper here means following emerging ABM methodology discussed later.) To do this one would have to collect and analyse data traditionally regarded as *both* qualitative and quantitative and the ABM would combine this distinctively. One can see this by recasting Schelling slightly to ask: 'How do individual residence change decisions lead to patterns of ethnic segregation?' Some data (about what neighbourhoods individuals thought 'acceptable' and how they decided to move) would have to be collected and analysed qualitatively (see, for example, Gladwin 1989).^[5] Other data would need to come from surveys and quantitative analysis: where different ethnic groups actually lived and what formal 'clustering properties' characterised this for example.^[6] Doubtless the opportunity to combine data thus

presents challenges but an argument would need to be made that these undermine the general principle that ABM integrate qualitative and quantitative data.

3.6 Thus the first major contribution of ABM is that it 'naturally' combines qualitative and quantitative data. Surprisingly, to my knowledge, previous work reporting ABM in sociology has not made this point. It appears that 'mixed methods' in sociology (Lopez-Fernandez & Molina-Azorin 2011 is typical) doesn't mean combining methods to develop and test the same theory but only using them within the same piece of research. Combining traditionally separated methods fully would be a major step forward for sociology and solid support for the (intuitive and widely expressed) view that qualitative and quantitative research are *both* necessary to effective social understanding.

Implication 2: Thinking differently about causes

4.1 Another preconception (coming partly from quantitative research and partly from the necessity of thinking about causality somehow or other) concerns how causes combine. Abbott (1988) has persuasively identified pitfalls in what he calls 'General Linear Reality'. This is the hope that quantitative methods actually mirror social reality so that, for example, 'small' causes have 'small' effects and causes 'add up' in stable ways. Even something as simple (and implausible) as Schelling's model may undermine this hope. This is shown using another advantage of ABM, the possibility for computational experiments. In particular, one can explore how system behaviour changes as agent PP changes.^[7] Before reading on, the reader might consider what they think will happen when a programme runs following rules describing Schelling's model for PP=0.6. The answer is shown in the centre of Figure 1. Perhaps not surprisingly agent type clusters form. Now consider two further questions. What is the *lowest* PP that produces clusters? What happens when PP=1? The answers are shown left and right in Figure 1 respectively. Less obviously, clusters form even when PP=0.3. This means that clustering occurs even when agents are 'content' to have a *minority* of own kind neighbours. This can reasonably be described as counter-intuitive. Typical theories of residential segregation (Charles 2003) suggest that individuals either 'choose' their own kind or deliberately exclude others. Here it is known (by construction) that neither effect operates and yet clustering still occurs. The article will return to the significance of this point shortly but the answer to the second question is even more striking. With PP=1 there is *no* clustering. Having seen clustering for PP=0.3 and PP=0.6, it is likely that most readers wouldn't have predicted this. It is clear that (in some obvious sense) the outcome with PP=0.6 is 'more clustered' than with PP=0.3. For example, clusters of different types are mostly non-adjoining and separated by empty site 'buffer zones'. Thus, the typical quantitative 'trend' with higher PP is (not surprisingly) towards 'more' clustering. But how many readers inferred this trend breaking down again before PP=1?

Figure 1. Typical simulation outcomes with PP=0.3, PP=0.6 and PP=1

4.2 To complete the picture, Figure 2 shows Percent Similar (a measure of clustering) and Percent Discontent (how many agents still wish to move) for values of PP 0-1 (increment 0.01).

Figure 2. Nonlinear effect of PP on clustering and content

4.3 This outcome is important to sociological debates about the micro-macro link (Alexander et al. 1987). Observing a 'macro' phenomenon (clustering), it is tempting to 'decompose it' (from the General Linear Reality perspective) into individual attributes that intuitively 'correspond' (strong preference for one's own kind and/or negative preference for others). Conversely, there is a temptation to 'add up' individual micro attributes (like increasing PP) to 'intuitive' macro patterns (like increasingly 'strong' clusters). Schelling's model shows, even for a phenomenally simple case (let alone presumably for cases of realistic social richness) that such informal inferences (from micro to macro or vice versa) can easily be mistaken.^[8] The reason is (following Abbott's argument) that causes typically do not add linearly. There comes a point where individual preferences simply cannot be satisfied within the system constraints. However, nothing in Schelling's assumptions (only in the dynamic process which most of us simply can't follow without the computer) signals this. Thus, registering the full implications of Schelling's model,^[9] a research method is urgently needed to build theory linking micro and macro explicitly (since doing this using linearity preconceptions creates considerable risk of error). As the article shows, ABM is such a method. These concerns are already recognised by quantitative and qualitative researchers. The macro to micro version is called the micro foundations problem (Pereira & Lima 1996). If a macroscopic regularity like stable relative social mobility across diverse nations is observed (Grusky & Hauser 1984) how can that be accounted for via individual level processes (job applications, promotions, firm bankruptcies) that presumably produce it causally? The micro to macro version is the generalisation problem (Payne & Williams 2005). Suppose ethnography is conducted on a firm. What aspects (which could be variables, practices and processes or any combination) might be identified *ex ante* to assert that the firm would succeed (or fail) within five years? Thus, far from inventing difficulties with existing practice, ABM suggests how recognised challenges in explicitly linking micro and macro processes (across both qualitative and quantitative sociology) might be met.

Implication 3: Reconsidering theory building

5.1 One can also see how theory in qualitative and quantitative research has come to mean something compatible with each method. In quantitative research it seems to mean statistical models themselves, and claims about social process are rarely made independently of models supposedly representing them. For this reason, statistical analysis itself tends not to develop theory. Instead, non-quantitative theories are translated into variables (see Sullivan 2001 for an example involving cultural capital). Thus while it might be plausible to theorise capital accumulation as a 'Matthew Effect' (Merton 1968), it isn't clear how typical statistical techniques in sociology would represent a process in which individual attributes changed over time as a result of social interactions as well as states of the system.^[10] What happens instead (as Abbott argues) is that models become synonymous with theory and 'explanations' tend to involve *ad hoc* process narratives not obviously inconsistent with observed regularities. (The potential hazards of this approach are demonstrated above.) On the other hand, qualitative theory often seems to involve narrative about the data itself or abstractions 'inspired by' the data rather than justified

generalisations from it.

5.2 This criticism is all very well (and necessarily can be little more than stereotyping for purposes of argument) but what *else* could theory mean? Again ABM suggests a significantly different answer. The Schelling model specifies a set of *processes* (a perceptual process identifying 'acceptable' neighbourhoods, a decision process – whether to move – and a migration process.) This is a complete specification of a simulated society *and its dynamics* regardless of whether Schelling's model plausibly represents any real society. This process specification 'theory' contains no 'abstract' theoretical concepts (latent functions, false consciousness), no causally/ontologically ambiguous statistical entities (market demand) and no 'conceptual gaps' (else the programme would crash). It is true that statistical entities can be derived from the model – for example the 'kind' of clusters occurring – but this is clearly different from these being built into it. Similarly, although PP might appear to be a 'theoretical' construct, it isn't. If PP comparison were *really* how people made migration decisions, respondent values would be accessible experimentally or perhaps even by self-report. It is only because it is so common for concepts like rationality to be *attributed* to individuals (rather than discovered from them) that it is hard not to assume that PP must be proxy for something respondents can't articulate. Unlike other approaches, if it became clear during ABM development that PP was *not* part of an actual decision process then the response would be to replace it by a more realistic (empirically based) process specification. (This point will be revisited later.) This is an example where existing methods (quantitative research) create a preconception that numerical elements in process specifications will be traditional 'factors' (like class) with associated conceptual problems. In fact, recast clearly in individual agency terms, there is nothing 'theoretical' about reporting: 'I save a fifth of my salary and spend the rest'. It is only when such claims are aggregated and regarded as macroscopically causal that their status becomes problematic.^[11]

5.3 Theory building can be well illustrated by reconsidering Schelling's model. Sociology recognises the simple (but nonetheless crucial) point that individuals 'act on' social structures *and vice versa*. This is the idea behind structuration (Giddens 1984). In a sense it is obviously true but neither qualitative nor quantitative research seem to transcend recognising its basic truth. What contribution can be made to this idea using ABM in general and Schelling's model in particular? Let us provisionally consider different agent type configurations as 'social structures'. If so, agent behaviour (movement) changes 'social structure' but social structure also changes agent behaviour. (A high PP agent moving somewhere surrounded by the other type will rapidly 'move on'. A low PP agent won't.) Although these 'structures' are infinitely malleable, no individual can remake them at will.^[12] (They can only cause others to leave and they do not intend this, it merely happens.) Thus, in Schelling's model (as with non-linearity) a challenging theoretical concept can be realised (at least in simple form) using a process representation absent ambiguities and hidden assumptions.

5.4 It is important to draw the right conclusions from this for my argument. This is *not* a model of what Giddens means by structuration. Motivating such a model, describing it and reporting results would undoubtedly capsize my stated intention to present more general ABM implications for sociological practice within the limited compass of a journal article. On the other hand, it seems implausible to claim that Schelling's model doesn't capture *something* about what structuration might (and perhaps even should) mean. It certainly shows a bi-directional interaction process between individuals and social entities ('neighbourhoods'). A rather abstract narrative theory has been instantiated as a fully specified process. However, it is possible to go further. An ABM can easily experiment computationally with different process specifications. It could be investigated when the system is 'determined' by agency (as when an agent knows enough never to move to empty sites where they will be discontent) or by structure (as when all possible moves for a discontented agent are currently blocked by house prices) or by interplay of both. It could be argued that whether or not this model does full justice to existing theoretical accounts of structuration, it may nonetheless form a more fruitful basis for empirical research (which to my knowledge existing structuration theory in its present form has not proved suitable for).

5.5 In preceding sections Schelling's model was used to illustrate how ABM offers three benefits addressing recognised sociological challenges: effective combination of qualitative and quantitative data, a useful framework for better understanding cause and a distinctive way of formulating theories as computationally implemented process descriptions (rather than equations and narratives associated with quantitative and qualitative research.) The rest of the article analyses emerging ABM methodology (less widely known than the technique itself) and considers its compatibility with sociological practice. This is done to show how it undermines various criticisms which might limit the sociological applicability of ABM.

Common objections to ABM and their relation to its methodology

6.1 While stressing that Schelling's model is only an example, it would be reasonable for readers to wonder whether the plausibility of arguments presented here depends on the example being so simple. It might be unclear how richer models should be developed and whether they would seem convincing compared to existing sociological approaches. The most effective answer involves presenting and discussing implications of the emerging ABM methodology.

6.2 This is very simply described (for fuller discussion, see Gilbert and Troitzsch 2005: 15–18).^[13] Defining a 'target system' (in Schelling's case, explaining urban ethnic segregation), a process specification is devised (given above) and implemented as a computer programme. (Different ways this may happen are discussed shortly.) The simulation is then run generating 'simulated data' (a distribution of ethnic residence). This is then compared with real data (using the same clustering measure on real neighbourhoods perhaps) and some assessment of similarity is made. (That assessment will also be revisited.) This process (subject to caveats below) underpins judgement about whether the process specification is 'adequate' or must be developed further.

6.3 Distinct ABM uses are characterised by how they generate process specifications and assess simulated and real data similarity. Schelling's model has already been presented as a 'toy' with assumptions instantiating 'sociological intuition' and not calibrated or validated.^[14] From the perspective of empirical sociology, however, a useful approach is to begin with a 'version 0' ABM systematically synthesising *existing* research on the target system. This identifies missing data, conceptual ambiguities and so on. Since it is rare that version 0 simulated and real data match, additional data gathering gives opportunities to explore and understand the ABM, refining it into 'version 1'. This process is repeated with successive versions each hopefully approximating social reality more closely. However, each version should try to maximise quality improvement because there are limited bases for comparison between simulated and real data. This is because, if the researcher redesigns Schelling's model to reproduce observed clustering, they can no longer claim clustering as evidence of model quality. Without systematic methodology, ABM faces the risk that unprincipled model adjustment may achieve fit on all available data by manipulation alone (rather than through achieving actual insight).

6.4 Given this methodology, researchers must make effective choices of target system and principled justifications for 'appropriate' similarity between simulated and real data. A system that is too simple (a straight line on a scatter diagram) cannot really justify ABM because many models (including much more parsimonious ones) can produce single straight lines.^[15] Conversely, it may be impossible to understand target systems that are too complicated (or have poorly defined boundaries) by *any* method. Obviously, the more similarity between simulated and real data (and for the most 'non-designed' features), the more likely it is that the ABM gives real insight into social reality (rather than simply summarising data). However, just as there is no 'right' (but only conventional) statistical significance level, how much similarity between real and simulated data is sufficient depends on simulation goals and on academic conventions. When ABM was even more novel (and even now for domains never previously analysed using ABM), any kind of 'non-designed' similarity is thought provoking. Now, however, the ABM community is starting to consider what constitutes 'adequate' model support. In Schelling's model, for example, it is found that the claim 'there are clusters' does not even distinguish variant model assumptions and is thus too weak to be useful. (It is now known how hard it is to build a Schelling style model that *doesn't* cluster.) A tougher model test would be reproducing 'adjacent' and 'buffered' clusters for different PP values in real neighbourhoods. A much tougher test would be (for three or more ethnic groups) explaining why clusters are separated or nested (like a target). An even tougher test would be (using only calibrated assumptions) mirroring actual residential segregation in a real town like Leicester or, toughest of all, predicting its future. But here the limits of what any existing research method can apparently do are reached.

6.5 Before considering common objections to ABM in the light of this methodology, it is useful to distinguish between empirical and technical limitations on theory building. Some approaches (like rational choice) make assumptions that are not empirical but 'necessary' to the application of the method.^[16] For example, if a theory requires analytic solutions to equations then equations must be chosen to admit analytic solutions regardless of whether these are plausible representations of social behaviour. The possibility of applying the method thus 'trumps' the plausibility of doing so. Another considerable advantage of ABM is thus that it removes all technical limitations because its models don't need to be 'solved' but merely unfolded dynamically. If one can conceptualise a process clearly, an ABM can be programmed. This immediately undermines assumptions (like perfect rationality) that can't be justified on empirical grounds alone. The defence 'if this isn't assumed, nothing can be done' loses force. ABM thus radically shifts the question in theory building from 'what is it necessary to assume?' to 'what is it possible to justify assuming empirically?'

6.6 As this article will now show, combining systematic methodology with 'deconstructing' technical limitations on theory building has dramatic implications for sociological practice.

Excessive simplicity

7.1 If, despite warnings repeatedly stressing that they shouldn't, readers take Schelling's model as an empirical claim about segregation rather than an intriguing example, they might plausibly claim it was 'too simple'. (For example, it is very hard to believe that house prices don't affect segregation.) However, there are three reasons why such objections don't count against the value of using ABM in sociology. The first concerns supporting evidence. Merely showing that *the world* is more complicated than ABM is not enough. As with debates between qualitative and quantitative researchers, those who draw attention to detail need to show not just that there *is* detail but that it *has effects* (in terms of quantitative outcomes for example.) It is not clear this can be done without simulation. Like statisticians, ABM researchers are fully aware that the world is more complex than models. Unlike them, proponents of ABM don't have to defend implausible assumptions for technical reasons. If a critic can show (or even suggest in a way that justifies the effort) why a process might have an effect it can be added to an ABM. The second reason concerns sociological goals. There is no practical reason why an ABM where every agent had a unique highly complicated cognitive process couldn't be built. However, the reason for *not* doing this isn't technical but scientific. Such an ABM would be useless (being as complicated as what it described) but more importantly (if it accurately reflected society), sociology itself would be pointless. Unless there are social regularities (and thus simplifications) of some kind to be found, sociology cannot be a social science. Thus (while not technically limited) ABM still needs to attend to the broadly agreed purposes of sociology, namely the search for empirical regularities (though it has the advantage of being able to integrate qualitative *and* quantitative regularities rather than trying to downplay one or the other). The third reason concerns ABM methodology. Because of non-linearity, it is not always possible to be confident that ethnographic inferences and statistical regularities reflect underlying social causes or effects.^[17] As such, despite matters of judgement (about how similar real and simulated data 'ought' to be for example),

it is not the case that existing data automatically trumps an ABM. It *might* turn out that a surprisingly simple ABM showed excellent fit with reality even if it did not reflect the preoccupations of ethnography (which, through lack of generalisation capacity were simply mistaken) or the patterns found in quantitative research (which through violation of statistical assumptions by non-linear systems were actually artefacts). Thus, although sociology is welcome to challenge the methodology of ABM, 'appropriate' model complexity must be established empirically and not merely using preconceptions of existing research methods.^[18]

Ad hoc models

8.1 It is necessary to distinguish objections to two kinds of *ad hoc* models. The first kind (which ABM formerly suffered from badly and still does somewhat despite emerging methodology and institutional regulation) is research without systematic procedure. Just as qualitative and quantitative research have been stereotyped for purposes of argument, ABM research can be stereotyped as reading a few pop science articles on social networks, translating these into an impromptu ABM and 'playing with it' until publishable results emerge. In such models there is no calibration or validation and no guarantee that a different modeller reading the same articles would even produce the same model. Such ABM don't develop effectively or compete to understand social reality but merely pile up. Clearly, this is unsatisfactory but is no more an argument against ABM generally than 'button pushing' statistics (by those not understanding the tests) or 'cherry picking' ethnography are arguments against those methods properly conducted. This argument involves consistent treatment: if badly conducted research doesn't undermine the value of existing methods when well conducted, then it is special pleading to argue that it undermines ABM.

8.2 The second objection is harder to answer because it often depends on implicit beliefs about how models 'ought' to work. For example, economists criticise models not producing unique equilibria and label them *ad hoc*. (See Morris and Shin 2001: 139.) However, it is reasonable to ask why economists believe real economies have unique equilibria so models should share this property. There are, inevitably, unsupported (though not implicit) beliefs in ABM (for example, that the better a properly constructed ABM can track real data, the more likely it is to have gained real insight into social process) and these must be subject to investigation and challenge but the issue of consistency of treatment also arises here. If existing research methods can define their own methodologies including unsupported beliefs, ABM is allowed to do so too.

8.3 As before, however, apart from general arguments about adequate evidence and consistent treatment, the distinctive nature of ABM also makes a practical contribution to avoiding *ad hoc* models. Because it uses programmes running on standard software, model sharing and replication are more common in ABM (see for example Will 2009) than in sociology generally and this opportunity to enforce good ABM practice is institutionally encouraged.

Parameters by another name

9.1 Since my argument questions the ontological and causal status of variables in quantitative analysis, readers might reasonably object that PP looks suspiciously like a variable too. However, as has already been suggested (but can now be argued more fully after methodological discussion) this apparent resemblance is mistaken. Leaving aside empirical plausibility (and how it arises), it is clear that in a particular behaviour ('save one fifth of your income and spend the rest') one fifth is not a parameter linking variables in the same sense as the (invented) claim 'women are one fifth more likely to reach university'. In the first case, the description (assuming it is honestly expressed and correctly recorded) directly explains agent behaviour (because that is how *they* decide on saving). In the second case, reporting an aggregate relationship (for a *sample* of men and women) may or may not directly explain anything. That differential could involve sexist selectors who are more likely to reject men or it could be a composite of many prior processes more or less approximated by model variables. Thus ABM gives us a perspective from which to unpick preconceptions (based on existing research methods) about what parameters are. Qualitative research rarely seems to credit (or analyse) decision-making processes containing quantitative elements. However, it cannot be assumed that social actors do not engage in decision making like this just because such assumptions are most often 'imposed on them' by theorists. If, empirically, people do make decisions involving numbers (or even equations) then our theory building tools must be able to represent this and trace out its implications. However, this must be done coherently in terms of ontology and causality. Even if individuals do decide using calculations, it is necessary to be very cautious (for reasons given above) in assuming that these will 'gross up' to statistical regularities (let alone robust descriptions) via macroscopic equations. Given previous discussion, it should be fairly clear why concerns about quantitative causality and ontology don't apply to ABM. Claims about cause and ontology are made directly in an ABM not mediated by equations. For example, in studying educational attainment by gender, an ABM must be clear when (for example) 'being a woman' matters because it affects how other agents treat you (sexism) and when it matters because you act from intrinsic differences (like earlier development or lower confidence). There are plenty of ways in which the variable 'gender' can (and undoubtedly does) affect educational outcomes but ABM (properly conducted) does not need to (and should not) elide them into variables and associated 'parameters'.

9.2 Summing up then, provided the use of quantitative elements in ABM is properly justified empirically (people actually do seem to make decisions using numbers in this context) and causally (this parameter affects the system because it is determining in the individual agent decision process) then it is not (despite surface appearances) subject to the concerns raised here about causality and ontology in statistics.

Mistaking drift for dogma (or weakness)

10.1 Being a critical advocate of ABM, one can certainly observe that some kinds of model are much more popular than others and that these aren't necessarily congenial to sociology. Social networks that are simply generated (like 'small worlds') are preferred to those with greater empirical plausibility. Models either tend to have very little 'geography' (like Schelling) or lots (with every river and road marked even when almost nothing is known about their bearing on social action). Very few ABM (except those specifically designed to represent organisations) involve agents acting contextually in social structures or institutions like 'home', 'work' or 'pub' (which most of us actually do nearly all the time).

10.2 However, it is important not to draw the wrong conclusions from this. As already suggested, this isn't occurring because ABM *cannot* build such models or because it considers them unimportant. It may be occurring (particularly with reference to existing fields like Social Network Analysis) because ABM researchers are not doing enough reading about what is already known (and perhaps institutional regulation needs to be stronger in this area). However, like any bad practice where good practice can be identified, this tendency doesn't undermine the approach in general. Instead, it requires us to find practical ways to promote good practice.

10.3 Based on this, the biggest pressure for ABM improvement in sociology is not going to be ignoring it and hoping it goes away (it seems to thrive on this though not necessarily in a positive way), nor raising general arguments against using the method at all (particularly if based on misunderstandings or implicit value judgments which simply cause hostility between groups) but by sociologists understanding the method well enough to challenge it on its own terms. It has been shown that ABM has a systematic methodology even if it isn't always followed. If an assumption is silly, data needs to be cited that shows it or research that suggests a more plausible process specification. If it is thought that the technique is ignoring a particular domain or kind of social behaviour, critics should build models themselves rather than grumbling that someone else hasn't done it. (Similarly, if it is thought that a neglected aspect of modelling 'matters', an ABM should be built showing how and why it does.) This process is likely to be very educational on both sides. ABM research will have to be much more careful about engaging with existing research traditions while sociologists will have to reflect carefully about when they really 'know' things (sufficient to convince others) rather than merely operating in a pleasant environment of shared belief. (From outside, the arguments economists use to justify rational choice as a theory of everything look pretty weak. From inside, serious alternatives appear almost unthinkable. It is more valuable scientifically to be able to convince your enemies than your friends.) A key goal of this article has been to lay out the issues clearly enough that such a mutual critical engagement might take place.

Conclusions

11.1 This article has shown, using simple examples throughout, how ABM can make important contributions in three areas of sociology (combining quantitative and qualitative data, thinking more effectively about causation and building effective theory). It has also shown how these contributions bear on recognised problems in sociological research (particularly how much detail 'matters' and how theorising should reconcile micro and macro phenomena). Finally, it has shown how, within the context of its emerging methodology, many commonplace objections to the approach do not (as believed) render ABM inapplicable in sociology.

11.2 Of necessity, a 'definitive' statement of ABM methodology is not presented since it is still emerging. Nor is it likely that this analysis will completely dissolve sociological reservations about ABM (although the intention is that issues surrounding these have been clarified thus allowing for productive dialogue). Clearly ABM is not a panacea and it is no harder to do bad simulation research than it is to do bad research of any other kind. Like any research method, ABM faces serious challenges already under consideration within its research community. For example, how are qualitative data 'translated' systematically into ABM? (See, for example, Agar 2003.) Nonetheless, it is hoped that this article has provided a clear analysis of ABM and its methodology and suggested why sociology, far from merely tolerating it, might find the approach actively helpful to long acknowledged sociological challenges.

Notes

¹ See for example Centola et al. (2005), Chattoe (2006), Gilbert (1997), Lomborg (1996), and Malleson et al. (2012).

² Since large-scale data is costly to collect, individuals are generally obliged to take what is available as given.

³ For example, see Bertaux and Thompson (1997) criticising quantitative social mobility research using qualitative data.

⁴ A theory like rational choice may be directly falsified by data (for example in the Ultimatum Game – Oosterbeek et al. 2004). Obviously a statistical model fitting data cannot fail to apply but instead demonstrates lack of real social insight by inability to predict out of sample, having non-robust parameters and so on.

⁵ It is only rather recently that ABM has considered converting qualitative data systematically into programme specifications. At this stage, it is only possible to refer interested readers to more-or-less imperfect examples (see Polhill et al. 2010 and Chattoe-Brown 2010 and the combination of interviews,

theory building and modelling discussed in Chattoe and Gilbert 1997, 1999, 2001) rather than say definitively how this is (or should be) done. For the kind of qualitative data that might be used as raw material for this process in the context of the Schelling model, see Winstanley et al. (2002) and Hickman (2010).

⁶ See, for example, Hatna and Benenson (2012).

⁷ The simulations here were written in a package called NetLogo. This is free and works on PC and Mac. It comes with a library of examples running directly on the web, see: <<http://ccl.northwestern.edu/netlogo/models/Segregation>>. This implementation of Schelling's model was developed by Uri Wilensky and is used for research purposes with his permission.

⁸ Space does not permit adequate discussion of the relationship between non-linearity, emergence and holism (reflecting on the hazards of reductionism) but these ideas are also of considerable sociological interest (Eve et al. 1997).

⁹ If a system this simple is non-linear how likely are real systems to be linear?

¹⁰ For example, a person might lose social capital as a result of failing to display appropriate cultural capital in interacting with a 'superior'. At the same time, inadequate cultural capital (relative to the aggregate level displayed by other 'competitors') might lead to their non-inclusion in an organisation, further weakening future opportunities for social capital development.

¹¹ This example is motivated by a paradigm case of mathematical modelling, namely the consumption function in economics which identifies a relationship between aggregate consumption and income (Molana 1993). The issue raised by ABM (through the Schelling model) is whether it is safe to assume that the aggregate consumption function might be decomposed into unique and stable individual behaviours or how assurance is achieved that individual behaviours empirically observed will add up to a stable aggregate function.

¹² There is nothing in Schelling's assumptions to stop *any* agent moving and thus any configuration being completely changed.

¹³ There seems to be an emerging consensus here. Epstein (1997) talks about 'generative' social science, the idea being that you have truly explained a particular pattern of data if you can reproduce it 'adequately' using a computational process specification. Axelrod (1997) talks about ABM as 'thought experiments' that sharpen our intuition about how particular empirical situations may have arisen.

¹⁴ However, as shown above, nonlinearity means this activity still generates thought provoking results.

¹⁵ It is good ABM design to focus on target systems with diverse research traditions giving raw material for both calibration and validation.

¹⁶ See, for example, the discussion of Leontief (1971).

¹⁷ This is really only a restatement of an earlier point. Using the Schelling model as an example, it turns out to be mistaken to assume that increasing individual 'intolerance' (rising PP) continues to produce increasing segregation (though in a linear system that would be a very reasonable inference). Conversely, it turns out to be mistaken to assume that clustering requires individual level intolerance (in that a PP of 0.3 – allowing for the majority of the other type as neighbours) – still produces it.

¹⁸ Similar arguments rebut accusations of 'excessive homogeneity' in ABM. If there is an empirical reason for thinking a system is non-homogeneous that can be incorporated into the ABM. However, qualitative researchers still have to suggest why heterogeneity might *matter* (rather than just showing it exists) and quantitative researchers cannot justify homogeneity assumptions on technical grounds when ABM simply doesn't need them.

References

ABBOTT, A. (1988) 'Transcending general linear reality', *Sociological Theory*, 6(1) p. 169–186.

AGAR, M. (2003) 'My Kingdom for a function: Modeling misadventures of the innumerate', *Journal of Artificial Societies and Social Simulation*, 6(3) <<http://jasss.soc.surrey.ac.uk/6/3/8.html>>.

ALEXANDER, J., Giesen, B., Münch, R. & Smelser, N. (Eds.) (1987) *The Micro-Macro Link*. Berkeley, CA: University of California Press.

AXELROD, R. (1997) 'Advancing the art of simulation in the social sciences', *Complexity*, 3(2) p. 16–22.

BARTHOLOMEW, D. (1973) *Stochastic Models for Social Processes*, second edition. London: Wiley.

BERTAUX, D. & Thompson, P. (Eds.) (1997) *Pathways to Social Class: A Qualitative Approach to Social Mobility*. Oxford: Clarendon Press.

- CENTOLA, D., Willer, R. & Macy, M. (2005) 'The Emperor's Dilemma: A computational model of self-enforcing norms', *American Journal of Sociology*, 110(4) p. 1009–1040.
- CHARLES, C. (2003) 'The dynamics of racial residential segregation', *Annual Review of Sociology*, 29 p. 167–207.
- CHATTOE, E. (2006) 'Using simulation to develop and test functionalist explanations: A case study of dynamic church membership', *British Journal of Sociology*, 57(3) p. 379–397.
- CHATTOE, E. & Gilbert, N. (1997) 'A simulation of adaptation mechanisms in budgetary decision making', in Conte, R., Hegselmann, R. & Terna, P. (Eds.) *Simulating Social Phenomena*. Berlin: Springer-Verlag.
- CHATTOE, E. & Gilbert, N. (1999) 'Talking about budgets: Time and uncertainty in household decision-making', *Sociology*, 33(1) p. 85–103.
- CHATTOE, E. & Gilbert, N. (2001) 'Understanding consumption: What interviews with retired households can reveal about budgetary decisions', *Sociological Research Online*, 6(3) <<http://www.socresonline.org.uk/6/3/chattoe.html>>.
- CHATTOE-BROWN, E. (2010) 'Building simulations systematically from published research: A sociological case study', draft paper, Department of Sociology, University of Leicester.
- EPSTEIN, J. (1997) *Generative Social Science: Studies in Agent-Based Computational Modelling*. Princeton, NJ: Princeton University Press.
- EVE, R., Horsfall, S. & Lee, M. (Eds.) (1997) *Chaos, Complexity, and Sociology: Myths, Models, and Theories*. London: Sage.
- FORRESTER, J. (1971) *World Dynamics*. Cambridge, MA: Wright-Allen Press.
- GIDDENS, A. (1984) *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge: Polity.
- GILBERT, N. (1997) 'A simulation of the structure of academic science', *Sociological Research Online*, 2(2) <<http://www.socresonline.org.uk/2/2/3.html>>.
- GILBERT, N. & Troitzsch, K. (2005) *Simulation for the Social Scientist*, second edition. Milton Keynes: Open University Press.
- GLADWIN, C. (1989) *Ethnographic Decision Tree Modelling*, Sage University Paper Series on Qualitative Research Methods Volume 19. London: Sage.
- GRUSKY, D. & Hauser, R. (1984) 'Comparative social mobility revisited: Models of convergence and divergence in 16 countries', *American Sociological Review*, 49(1) p. 19–38.
- HATNA, E. & Benenson, I. (2012) 'The Schelling model of ethnic residential dynamics: Beyond the integrated – segregated dichotomy of patterns', *Journal of Artificial Societies and Social Simulation*, 15(1) <<http://jasss.soc.surrey.ac.uk/15/1/6.html>>.
- HICKMAN, P. (2010) 'Understanding residential mobility and immobility in challenging neighbourhoods', Research Paper Number 8, Centre for Regional Economic and Social Research, Sheffield Hallam University, September.
- LEONTIEF, W. (1971) 'Theoretical assumptions and nonobservable facts', *American Economic Review*, 61(1) p. 1–7.
- LOMBORG, B. (1996) 'Nucleus and shield: The evolution of social structure in the Iterated Prisoner's Dilemma', *American Sociological Review*, 61(2) p. 278–307.
- LOPEZ-FERNANDEZ, O. & Molina-Azorin, J. (2011) 'The use of mixed methods research in the field of behavioural sciences', *Quality and Quantity*, 45(6) p. 1459–1472.
- MALLESON, N., See, L., Evans, E. & Heppenstall, A. (2012) 'Implementing comprehensive offender behaviour in a realistic agent-based model of burglary', *Simulation*, Vol. 88, No. 1, pp. 50–71.
- MOLANA, H. (1993) 'The role of income in the consumption function: A review of on-going developments', *Scottish Journal of Political Economy*, 40(3) p. 335–352.
- MORRIS, S. & Shin, H. (2001) 'Rethinking multiple equilibria in macroeconomic modeling', in Bernanke, B. & Rogoff, K. (Eds.) *NBER Macroeconomics Annual 2000* (Vol. 15). Cambridge, MA: MIT Press.
- MERTON, R. (1968) 'The Matthew Effect in science', *Science*, 159(3810) p. 56–63.
- OOSTERBEEK, H., Sloof, R. & Van De Kuilen, G. (2004) 'Cultural differences in Ultimatum Game experiments: Evidence from a meta-analysis', *Experimental Economics*, 7(2) p. 171–188.
- PAYNE, G. & Williams, M. (2005) 'Generalization in qualitative research', *Sociology*, 39(2) p. 295–314.

PEREIRA, L. & Lima, G. (1996) 'The irreducibility of macro to microeconomics: A methodological approach', *Revista de Economia Política*, 16(3) p. 15–39.

POLHILL, J., Sutherland, L.-A. & Gotts, N. (2010) 'Using qualitative evidence to enhance an agent-based modelling system for studying land use change', *Journal of Artificial Societies and Social Simulation*, 13(2) <<http://jasss.soc.surrey.ac.uk/13/2/10.html>>.

SCHELLING, T. (1969) 'Models of segregation', *American Economic Review*, 59(2) p. 488–493.

SULLIVAN, A. (2001) 'Cultural capital and educational attainment', *Sociology*, 35(4) p. 893–912

WILL, O. (2009) 'Resolving a replication that failed: News on the Macy and Sato model', *Journal of Artificial Societies and Social Simulation*, 12(4) <<http://jasss.soc.surrey.ac.uk/12/4/11.html>>.

WINSTANLEY, A, Thorns, D. & Perkins, H. (2002) 'Moving house, creating home: Exploring residential mobility', *Housing Studies*, 16(6) p. 813–832.