PART C: AGENT-BASED MODELS

6.0 AGENT-BASED MODELLING AS THE METHOD OF ANALYTIC SOCIOLOGY

6.1 AS and Agent-Based Models

One, but only one, of the critical challenges to AS is to trace through the concatenations of individual lines of activity to demonstrate how the macro emerges from the micro. Unless this can be done, AS explanations will be just as black box-like as other sociologies and the accounts given little more than causal stories. Such traceability is one of the central requirements of mechanism explanations. Agent-Based Models (ABM), or so it is thought, provide AS with the opportunity to address this challenge.

> *Although systematic regularities can be observed at the (macro) level, such associations typically say little about why we observe what we observe. From this perspective such associations are surface phenomena arising from and requiring explanations in terms of deeper underlying processes. The important point is that at each instance of time, macro, … is supervenient on micro.... and in order to explain changes in (the macro) we must explain changes in this micro base. Until very recently we did not have the analytical tools needed for analyzing the dynamics of complex systems that large groups of interacting individuals represent. Powerful computers and simulation software have changed the picture—so much so that it is possible to have some confidence that agent-based computer simulations will transform important parts of sociological theory because they allow for rigorous theoretical analyses of large and complex social systems. (Headstrom and Bearman 2009 p12)*

This is an important claim. Robert Merton was quite convinced that what was holding sociology back was its immaturity as a scientific discipline. In Richard Feynman's phrase, sociology "hasn't done the work" to be able to offer the empirically based general theories that other disciplines can. For AS, at least, it seems that ABM will allow that work to be done. ABM will close the explanatory gap, resolve the mystery of supervenience and provide a method for robust social theory generation. If this is to be so, then the ability of ABM not just to bridge the micro-macro divide but to do so in a way that can be grounded in data collected on the detail of the lives and activities of ordinary individuals must be demonstrated. If ABM cannot do both of these things, then, adding ABM to AS will add only wishful thinking to what already is largely an exercise in future perfect aspiration.

In the following discussion, we will examine this challenge. First, we focus on the problems which users of any (in this case computational) model have to overcome in order to apply that model to their own research problem. Following this, we take up the methods by which theory is formalised in a model and the relative advantages and disadvantages of formalisation itself. These are then illustrated by reference to an actual model, the simulation of cultural theory constructed by Mercedes Bleda and Simon Shackley (2012)

The iron law of unintended consequences gives sociology its topics. Sitting somewhere between Murphy's Law ("If something can go wrong, it will go wrong") and Newton's Third Law of Motion ("To every action, there is an equal and opposite reaction"), the iron law of unintended consequences tells us that any social action will have unforeseen, unintended consequences, many of which will, in all likelihood, be deleterious. Testimony to the operation of the iron law can be found both in commonsense ("it seemed like a good idea at the time") and academic disciplines. The infamous designation of the problems of the policy sciences as 'wicked' captures the impossibility of predicting and managing all the consequences of well intentioned policy interventions. Where sociology is distinctive among these policy sciences is that it usually sees one of its task to be that of the identification of whose interests are promoted by the unintended consequences that ensued and, hence, who might be held responsible for them.

Of course, since sociology is itself social action, it too is subject to the iron law. Undertaking sociological investigations, developing sociological theories, analysing sociological data are all social activities and, inevitably, have unintended consequences. In this discussion, we want to explore one set of such consequences, those which follow from the use of Agent-Based Modelling (ABM) techniques to provide a more scientific or otherwise enhanced basis to sociological theorising. Our aim will be to bring out how the desire to cast theory into rigorous, formal models leads paradoxically to operationalise models which are less than transparent, predictable and consistent. To paraphrase Harold Garfinkel, we illustrate how good scientific practice might lead to poor sociological theories.

6.2 The Rationale for ABM

Simulation, and to a lesser extent ABM, have been used in sociology for quite some time. Their proponents argue for them on the grounds they either solve or avoid many of those disciplines' perennial problems.[19] One leading light, Robert Axelrod (2006), suggests that, unlike other forms of simulation, ABM's promise is far greater than simply being a facility for undertaking precisely defined *gedanken experiments*. It offers no less than a new and third way of doing science, one that is on a par with the inductive experimentalism and deductive reasoning of the natural and mathematical sciences. Not only does this new form of science offer a way for the sociology to escape from its self-imposed inferiority complex and hence cease attempts to parody the natural sciences (as Stanley Lieberson put it), it also offers a robust platform for novel interdisciplinary work across the social and natural sciences. As proof of this, Axelrod points to the results of his work with the

---

[19] For an introduction to ABM see Gilbert (2008). For a survey of types of ABM models see An (2012). Perhaps the most well knowm social science ABMs are Alexrod's co-operative prisoners' dilemma (1997), Reynolds' "boids" (1987) and Dean et al's study of the Anasazi (1999). For a far more sceptical view of simulation, see Frigg & Reiss (2009)

evolutionary biologist William Hamilton. Using ABM, Axelrod and Hamilton 'proved' Hamilton's conjecture that it was the need to manage infestation by parasites which led to the development of sexual (as opposed to asexual) reproduction in early animals.

ABM is a form of *microsimulation*. Agents are individual 'actors' modeled with particular attributes and motivated to follow specific decision rules within a defined 'environment'. This environment is the *virtual world* of the simulation. ABM adopts strong methodological individualism with a 'bottom up' approach to modeling social structures and processes rather than the 'top down' approach associated with structural models. As we have seen, this is its major attraction for AS. It allows the observation of structural phenomena as *emergent* consequences of individual's actions.[20]

A second advantage of ABM is felt to be its capacity to deal with multidimensionality, something which other forms of modeling are said to have trouble with.

> *Traditional formal modeling methods rely on algebraic formulas to translate simple relationships into mathematical expressions. But the limits of algebra prevent such techniques from incorporating many of the things known to be true of most of the worlds social scientists find interesting, including their multidimensionality, the presence of large numbers of interacting and autonomous units, and the predominance of highly irregular but nonrandomized patterns in the distribution of traits or interaction styles. The constraint of algebraic solvability therefore limits the ability of traditional formal modeling methods to capture the richness of interesting and even well-established substantive theories. Reliance only on traditional formal modeling techniques (such as game theory and rational choice) thus often entails ignoring what the modelers actually believe to be true about analytically crucial parts of the world. However, ABM, and the computer-assisted bottom-up simulations it produces, can be designed to capture beliefs about the real world embedded in or expressed by good substantive theories, thereby providing researchers with new opportunities to examine possible and probable outcomes associated with specific theory-based claims. (Lustick & Miodownik 2009 p 223)*

By running a model a number of times, we can generate what might be thought of as 'quasi-empirical data' which can be compared to data derived from more standard investigative methods such as published official data, surveys, ethnographies and interviews. This aspect of ABM is likened by its proponents to the 'thought experiments' beloved of social scientists and the experimental runs of the natural scientist.

Like structural models, ABM models are *formal*. They are computational models which represent theorised social actors and social structures. The models take inputs which are values for the assumed 'drivers' of the model and generate outputs which are the resulting states and structures. The formality of the modelling is an important feature of ABM. The rules of computability place constraints on what can be expressed in and what can be done with a model in exactly the same way that other forms of mathematical logic constrain structural models.

---

[20] For an attempt to re-formulate AS so that it can relax the stipulation of methodological individualism see Marchioni and Ylikowski (2013). If successful, this will have implications for the specification of ABM in the service of AS.

It is also important to remember that ABM models are not *explanatory* in the sense that structural models aim to be. The degree of mapping between the behaviour of the agent in the model and actual social actors is problematic. Rather, the aim of an ABM is to act as a formalised "idea pump", clarifying internal relationships within the model and by providing alternative future 'possible worlds'.[21] As we will see, these characteristics make the problem of re-deployment one of ABMs central challenges; a fact that is widely acknowledged by its proponents.

Most sociological theories are discursive in form and generalise from a single, or at best a small number of cases. As a consequence, even if they are framed in terms of a few key variables, the need to ensure that the theory accommodates the complexities of the case from which it is derived leads to the theories themselves becoming complex. This complexity makes it difficult to demonstrate broad generalisability across a variety of conditions. Putting it somewhat differently, sociological theories tend to be constructed for particular sociological phenomena and hence defy detailed replication. It is commonly agreed this is just another symptom of the discipline's immaturity.

Discussing this generalisability issue in the context of complex social situations, H.M. Blalock (1979) observed two different tendencies. The first is to provide extensive and elaborate detail for each modeled case. While this does increase the realism and empirical reference of the cases, it makes it difficult to determine the degree of comparability across cases. The continuities are drowned out in the volume of contextual detail. The second tendency is to allow the array of available data to determine the scope of variables in the model. Thus, if there is no available data for some measure to be meaningfully measured, the variable is dropped from the model either because it can be deemed immaterial or because it has been absorbed into other variables.

> *One sees in the journal literature many path diagrams involving six or eight variables, whereas even a minimally realistic theory would undoubtedly require twenty or thirty......the tendency to omit variables from one's theory on the grounds of empirical expediency remains a serious problem for the discipline. (Blalock 1979 p. 130)*

As a consequence, working back from the model to the situation on which it is based, the user of any model has somehow to accomplish an elision between the situation as described and the situation as modeled.

Two strategies of abstraction have been widely recommended to overcome this limitation: formalisation and modelling. It is their use which generates the difficulties we will outline. Formalisation is about the mode of expression. When a theory is re-written in formal terms, it is cast into an axiomatised abstract language which is constrained by the nature of its grammar; that is, propositions in that language are derived as what are self-evident truths (tautologies) from defined postulates and the axioms and can only be manipulated in certain ways. Two closely related modes of formalisation are mathematics and logic. Analysis of formalisation consists in working through the set of interconnected tautologies which constitute the

---

[21] This is an important point we will come back to. Models are purposeful. if you like, they are just as interest-relative as explanations. The criteria for selecting the features to be modelled are functions of the modellers ambitions for the model. See (Giere 2010)

framework. The deductions made within that framework are themselves tautologies Analysis consists in working through the set of interconnected tautologies. Thus unknown and unanticipated 'truths' about the model can be discovered/uncovered through analysis.

Models are purposeful simplifications (Giere 2010). The grammar of formal models provides sets of defined rules to secure structural symmetry between the 'target' phenomenon and its model. Formal models are concerned only with features where structural symmetry can be preserved. Informal models are less constrained. With these, simplification is achieved by prioritising a sub-set of variables which is taken to reflect a similarity to be examined. In Informal models this similarity rests on metaphor. In formal models, it rests on analogy. A classic example of an informal model in science is the 'planetary' description of the structure of the atom. The model provides nothing other than an image of the spatial arrangement of the particles which make up the atom. An informal model cannot reveal new truths.

6.3 Normal Natural Troubles of ABM

Good models have a number of properties. They should be:

- Simple: The model should contain the fewest possible assumptions. Complexity in models allows them to be manipulated to 'save the appearances' of any set of data to which they might be applied.
- Tractable: The model should be easy to analyse. Ease of analysis refers to the level of analytic resource required 'to run' the model. A model which can be drawn on a board or worked through on the back of an envelope is more tractable than one which can only be summarised in a description as long as War and Peace or resolved using supercomputing.
- Insightful: Since models are about concepts, they should be conceptually insightful. They should reveal fundamental properties of the phenomenon being modelled. Because a false model might have some conceptual value (one classic justification for rational actor theory in economics), an emphasis on conceptual insightfulness might lead one to give reduced weight to empirical reference.
- Generalisable: A model which is only applicable to a single case is no more than a simplified re-description. A generalisable model should be applicable to a range of cases though not necessarily universal.
- Testable: A testable model enables predictions or projections which provide assessment of its validity. The tighter or more precise the predictions which go 'beyond the data given', the stronger the test of the model.
- Empirically referential: To be testable the model must map onto the world in some way. This mapping will be specified in the form of the modelling and is where structural symmetry

becomes important. This is a critical element for AS. Unless ABM models can be mapped onto actual social situations, the robustness of the model cannot be guaranteed.

Although it is possible to give crisp descriptions of the above criteria, all are essentially relative. Fixing the extent to which any one of them has been satisfied is a matter of judgment. Second, not surprisingly, building a model to satisfy all the above is extremely difficult. Researchers are usually forced to place greater value on some rather than others and so seek to find a balance which *satisfices* rather than optimises across the set. Because they only have recourse to the published model (and sometimes, in the case of computational models, the published code), the main difficulty in re-using someone else's model is uncovering the choices made. Although some are listed in the text or documentation; often they are not. In any event, they are usually unavailable in the detail required by anyone who wants to apply a model to a new setting.

As is well known, empirical reference is the central problem for ABM. Breen (2009) distinguishes two elements; adequacy and plausibility of reference. ABM researchers acknowledge this issue and expressly seek an acceptable degree of realism for their models. The usual approach is to ground the model on the consensus of the research literature of the specified domain. However, specification and implementation of this consensus rest on a number of simplifying assumptions or decisions. The complexity of the social world is reduced to make modelling manageable and the mathematics tractable. Thus determining the trade-off between realism and simplicity in any model is a first and, in the eyes of Midgley and his colleagues (2007) possibly the most difficult, issue to tie down. What modellers *cannot* do in their listing of assumptions embedded in their model is lay out all the ramified problematic possibilities which might have been included in the model but were not. Experts in the domain being modelled might be able to guess what else the literature covers which might have been included, but without a calibration of the literature-as-modelled with the literature-as-read, there will still be uncertainty (See Giere (op cit)). Those who are not experts in the domain can only surmise that some things *must have been left out* but precisely what and precisely with what consequences remains unknown. The user of the model has then to accept a fair degree of uncertainty as to the veridicality of the representations used.

A second very general trade off is that between a model's verification and its validity. A simple formulation summarises what is at issue here: verification is the assurance that the equations in the model run correctly; validation is the assurance that the correct equations have been run. Whilst modellers aspire to do both, often the economics of effort encourages them to focus on the former rather than the latter. Failure to implement the model correctly leads to *errors*. Failure to understand the significance of the additional *non-formalised* assumptions made to get the program to work leads to *artefacts*. Incorrect implementation leads to a failure of verification. Insufficiently well understood assumptions lead to a failure of validity. Since getting the code stable and running in a predictable fashion is a pre-requisite to having a model to describe and publish, it is hardly surprising that effort is skewed towards verification. The verification/validity problem can be construed as a Type 1 and Type 2 error over causality. Given the lack of clarity about verification and validity, precisely where the equilibrium point might be between too broad and too narrow a circumscription of causal processes in a model is often unknown. Too broad a cluster and many immaterial 'causes' might be included. Too narrow a cluster and some key ones might be excluded.

The origin of errors and artefacts lies in the division of labour necessary to turn a schematic model into an executable program. Galán et al (2009) separate out the following roles: thematician, modeler, computer scientist, programmer. Of course, all these roles could be carried out by the same person but more usually it is by a team. Apart from that of the programmer, the task of the roles is to produce a set of specifications from the design which is passed to them. Thus the thematician turns a discursive and often idealised description into a formalisable structure. The modeller turns the thematician's specifications into a formal model; and so on. At each transition, the specifications are re-written in a new format. Such translation introduces the possibility of error and artefact, but, unless a formal process of symmetry testing is carried out, exactly where will once more remain unknown. The irony is that it is precisely because the evidence to test symmetry of this kind is unavailable to sociology that AS has turned to ABM.

Errors are introduced when the re-specification process includes a misinterpretation or mistake with regard to the intentions behind the design. An obvious example is when the design asks for some computation to be looped through a population of agents but the programmer implements the loop for only a subset. Less easy to spot without access to the design documentation are examples such as the presumption that some calculating engine (say, a random number generator) coded in the simulation utilises real arithmetic when in fact it uses floating point arithmetic. Even more difficult to uncover are the consequences of inconsistent garbage collection leading to memory faults. Both of these can distort the way the system runs and the outputs generated, but in ways that remain unknown.

Errors are not always revealed by eccentric behaviour on the part of the executable code. Some 'bugs' are never discovered or only reveal themselves when amendments and extensions are made to the code base. Here again another trade-off is left obscure, that between 'hacking the model' and 'engineering the model'. With the latter, use is made of the many practices used by software engineering to ensure the robustness of the code base. Processes such as public code walkthroughs, detailed documentation and regular builds, smoke tests and so on, help to reduce errors and inconsistencies that otherwise accumulate during development and which can cause the model to be unstable and crash. These practices require time, effort, and perhaps most importantly, well trained and reasonably large and experienced project teams. None of these are in plentiful supply in most ABM projects. As a consequence, hacking the code by means of *ad hoc* fixes during build and early testing becomes the only other way to ensure code stability. However, this can lead to code structures that are either impossible to understand or impossible to predict (or both).

Other methods of securing verification come with their own troubles. Sensitivity tests can be run on the data generated by the model to determine whether the outputs retain the plausibility of the original specification. Again, however, the selection of which tests to run and over which variables and for how long is a matter of judgment. The failure of the model to "fall over" or generate extreme results is often taken as sufficient evidence of its verification with the outputs generated being summarized in rationalised histories of activity (or if we want to be unkind, Just So Stories). A more ambitious (and hence very costly) verification technique is "docking" (Axtell et al 1996). This is a reimplementation of the model in an alternative language using an alternative simulation tool. The aim of "docking" is to achieve a process akin to the *experimental replication of results* required in the mathematical and natural sciences by elimination of contextual influences of programming language, simulation tool and so on. Not surprisingly, given the difficulty of getting just one version of the simulation to run and the practical constraints on research time and other resources, utilising testing through docking as a general practice remains largely an ambition. In any event, neither sensitivity testing nor docking provide formal proof of the verification of the program.

Midgeley et al recommend using another technique which is potentially even more costly and risky, namely Miller's 'optimisation to destruction'. This uses a genetic algorithm to optimise perturbations in the model's performance by re-valuing a selected set of parameters. This method can show quite quickly just how unstable (ie likely to lead to extreme results if marginally changed) some parameters might be. Also, it can illuminate faulty code. The same approach can be used in terms of fitting output data of various kinds to actual empirical data. The problem is that any model will fail under extreme testing and so the risk is not that the model will fail but that it will fail well within what were thought to be its safe limits. Naturally, few researchers are willing to run such risks and subject what they think is a publishable model to stress testing of this kind.

It is important to understand that verification is not simply (!) a matter of exhaustively hunting down errors. As Midgley et al point out, there is no consensus in the computer science community that a program (ie complex model) can be fully proved to be error free. Therefore, although there is much that can be done to provide as high a level of assurance as is practicable, it remains for the modeller to determine the limits to verification.

Artefacts are a result of a mismatch between the assumptions which are thought to be producing the model's behaviour and those factors which actually are producing it. Here, following Galán et al, we can distinguish *core assumptions* from *accessory* ones. Accessory assumptions are not believed to be drivers of the model but required simply to make it work. However, when the model is run it may be impossible to tell if some outcome is a product of a core driver or of some accessory assumption. That is to say, what was felt to be an insignificant design decision may have turned out to have a significant impact. Decisions over topology of the environment or scaling systems for variables might be felt to be significant and yet, when the model is run, can turn out to have major implications. For Galán et al, it is usually artefacts which create problems of empirical reference.

> *.....the challenge of understanding a computer simulation does not end when one has eliminated any "errors". The difficult task of identifying what parts of the code are generating a particular set of outputs remains. Stated differently, this is the challenge of discovering which assumptions in the model are responsible for the aspects of the results we consider significant. Thus, a substantial part of this non-trivial task consists in detecting and avoiding artefacts - significant phenomena*

> *caused by accessory assumptions in the model that are (mistakenly) deemed irrelevant to the significant results.(Op Cit para 1.5)*

Another trade-off is between *face validity* and *construct validity*. For face validity, all a model has to display is informal similarity (as with the planetary model of the atom discussed earlier). To achieve construct validity, formal similarity based upon rigorous calibration will be required. For ABM, this could happen on two levels: the micro level of individual agent behaviour - a mapping of the model onto actual behaviour of the relevant real world actors; and the macro level where issue is the degree of goodness of fit of emergent structures. For the latter, the mapping should be over a reasonably lengthy period of time (the 'turns' in the model generally being much faster than those in the world). Because of the difficulties of mapping at both levels, what counts as sufficient evidence to provide an estimation of goodness of fit remains the judgment of the modeller. As a result such estimations tend to result in a 'good enough' rather than a best fit. As we suggested a few moments ago, the payoff of ABM for AS will depend entirely on the strength of the goodness of fit simply because sociology cannot provide good enough tests for itself.

Empirical reference in simulation depends upon three different but highly connected things: the degree to which the simplification required for modelling distorts the model's realism; the level of verification of the code base of the implemented computational model; the validation of the 'behaviour' of the model at the micro and macro levels. As things currently stand, all three components are known to be uncertain to some degree. As a consequence, simulations often displays classic goal displacement. The model building tends to take on a life of its own rather than being in the service of empirical investigation. In Midgley et al's words:

> *The major conclusion from our efforts to develop an empirical validation methodology is that we need to be much more influenced by the type and nature of the data we can plausibly obtain* ***before*** *we begin to specify our AB models, rather than developing from theory and then seeking appropriate data to fit the demands of this theory. (Midgeley et al op cit p 21 emphasis in original)*

The 'troubles' we have just sketched will be of no news at all to active simulation modelers. They know these things and bear them in mind when they interpret models published in the literature, models sent to them for analysis and comment, and models they want to 'tweak' and re-use. Since published papers, like all documents, are written to be read, authors presume an understanding of the above troubles as part of the readers' competences. This provides a practical solution to the problems of inclusion and exclusion of topics and detail. When choices are made over what readers will need to have explained and in how much detail, the generally known 'troubles' of using models can be ignored. As a result the presentation of the model in the published paper is reflexive on assumptions about the competences of the readership. Although an author might explicitly orient to a course of action type such as 'a-reader-with-a-working-knowledge-of-simulation' in the construction of the publication, this type will not cover every possible reader/user of the model. Somehow, the author has to decide where the line defining relevant explanatory detail should be drawn. Inevitably, some readers will fall the other side of the line. The author has to decide just what detail should be documented in the model since trying to cover all possible readers/uses would mean endless provision of ramifying detail. The author's problem is writing the model for a readership; the reader's problem is determining how that readership has been constructed ("Am I a reader for this model? How do I resolve its indexicality?").

Our argument in this section has been that attempting to adhere to the normative character of the virtues of scientific models leads simulation to a number of well-known but not often discussed troubles. These are not the product of incompetence, inattention and dereliction. They arise simply because of the trade-offs made in building and executing models and in representing a model in publishable form. It is important to recognise that though these troubles might take a particular character in ABM, they are not peculiar to it. All forms of modelling created for one group of users and re-used by others have the same troubles. In so far as it is done and however it is done, the task of writing and reading models is the task of resolving these troubles.

Having identified these 'troubles' and suggested their importance for AS' use of simulation, we now turn to a particular ABM simulation to trace through the processes by which an informal, discursive sociology theory is translated into a formal model.

7.0 CULTURAL THEORY AND FORMALISATION

7.1 Formalising Sociological theory

Let us recap. AS has turned to ABM because more traditional sociological methods do not focus on the fine grained patterns of individual actions which will allow middle range theorising of the linkage between micro and macro social structures to be built 'from the bottom up' and ABM does (or so it is said). ABM techniques should, therefore, permit AS to trace causality and supervenience *in silico* if not *in vivo*. Since ABM is a type of computational modeling, it is formal in character. Formalisation and modeling each bring a number of associated problems. In the previous section we discussed some of the latter. To focus our discussion of the former, we will look at a particular instance of the formalisation of a sociological theory. Although ABM and other types of simulation have been applied to social settings, there have been few which have actually tried to simulate how a specific sociological theory might apply to a particular event or set of events in a defined setting. Because they try this and do so in a clear and systematic way, even though it is not actually an ABM, we have chosen Mercedes Bleda and Simon Shackley's (2012) model, COWCULT, as our example. Bleda and Shackley are very clear that their discussion should be taken as a first foray into the use of simulation in relation to Cultural Theory and not as a final demonstration of its efficacy and so, in looking at COWCULT, our concern will not be with how well the model simulates the actual setting to which it is applied nor with how much support it gives to the theory it selects, Cultural Theory (CT), as an explanation of the events to which it relates. We will simply be concerned with the processes that have to be gone through to turn a discursive, informal sociological theory into a formal simulation model. By using simulation and especially ABM, AS wishes to avoid a number of methodological difficulties. We want to see if, in wishing to escape the frying pan, they might be likely to land in the fire.

7.2 Cultural Theory: the background

Cultural Theory is not a distinct theory in sociology but rather a myriad of broad approaches to understanding the social organisation of knowledge (and cognition in general). In many ways, you could say it *is* the social theory of knowledge in as much as its basic tenet is that knowledge is socially organised. It is then to be contrasted with non-social (such as psychological) theories. Formalising Cultural Theory is not so much formalising a theory within the sociology of knowledge but formalising the sociology of knowledge itself. Since in the case that we will discuss, Bleda and Shackley (op. cit.) refer throughout to Cultural Theory (CT) rather than the sociology of knowledge, we will follow their usage.

It is also important to remember that the sociology of knowledge is as much concerned with 'ordinary' or 'commonsense' knowledge as it is with the highly advanced knowledge structures of the sciences and related disciplines. Its claim (which we will not debate here) is that the knowledge articulated by the sciences is just as much socially organised as is ordinary or commonsense knowledge.

The underlying premise of CT is that we do not see 'things', we see them *as* things. This is what is called "Brentano intentionality". Perception involves the deployment of classificatory schemas, and it is these schemas which are socially organised. So, to see this as a ball, this as a car and this as a cow, we need to have the appropriate schemas in which there are balls, cars and cows and their distinctive properties. The social relativity of such schemas is the staple of much Cognitive Anthropology, where study after study has demonstrated that members of other societies do not see what 'we' take as obvious and natural distinctions and similarities.

The debates in CT are over how to account for the origins and consequences of the social organisation of knowledge and understanding. Once again, a central assumption is that there will be a functional fit between social structure and the social organisation of knowledge. This is not a hypothesis or conjecture. It is the starting point for CT.

Although there are differences in vocabulary (and hence of connotations), the narrative of all CT is the much the same. Forms of knowledge (and hence of perception) and the structure of society are mutually interacting force fields. Each shapes or resonates with the other. There is a reciprocal relationship between them and neither has causal priority. In Weber's terms, they have an "elective affinity" in which each is moulded by the other. Numerous studies of this affinity have been undertaken; possessive individualism and Puritanism; the engineering of clockwork and the mechanical universe of 17th century physics; or in our own time, computer engineering and 'information models' of the brain, society and so on.

Mary Douglas and Aaron Wildavsky (1983) have taken this set of premises and the tenets based on them and applied them to our understanding (or perception) of risk - that is, to what we see as being actually or potentially risky. At its core is the theory of grid/group co-variation of types of social structure and sets of beliefs or values which Douglas developed over several years (Douglas 1970). As with all sociology of knowledge, this theory suggests that belief and values are correlated with types of social organisation.

> *Group is the outside boundary that people have erected between themselves and the outside world. Grid means all the other social distinctions and delegations of authority that they use to limit how people behave to one another. A society organised by hierarchy would have many group-encircling and group-identifying regulations plus many grid constraints on how to act. An individualist society would leave to individuals maximum freedom to negotiate with each other, so it would have no effective group boundaries and no insulating constraints on private dealings. A sectarian society would be recognizable by strong barriers identifying and separating the community from non-members, but it would be so egalitarian that it would have no leaders and no rules of precedence or protocol telling people how to behave...... (Douglas & Wildavsky 1982 pp138-9)*

The types (hierarchical, individualist, egalitarian, sectarian) are generated by the binary forces of grid and group.[22] Such forces are either strong or weak. This 2 x 2 theoretical formulation produces 4 possible types of social grouping. The propositional attitudes expressed in beliefs and values of these types will vary according to the context set by grid/group interactions. To understand the particular set of beliefs or values held by any one type, it is necessary to understand the complexities of their grid/group interactions. This understanding is provided by the ethnographic or other evidence which the sociologist gathers on the functioning social organisation. What that understanding will provide is an appreciation of the ways that social organisation and cultural forms such as beliefs, values and attitudes towards risk are *mutually elaborating and legitimating*.

7.3 The Formalisation Task

The Douglas and Wildavsky CT is a discursive functional narrative. Bleda and Shackley set out to translate it into a formal model which can then be used to simulate the development of beliefs, attitudes and values. The setting they apply it to is the UK’s BSE crisis in the late 1980s and the perceptions of risk that arose during it. However, while they use the BSE case, their model would not be a formal if it could only be used to model BSE. As we will discuss later, formal models aim to be context free.

A number of steps have to be taken to build a formal model from CT.

1. A formalisable description of the functional base of CT has to be provided. This will require two sub-steps. First, CT has to be re-cast as a componentised theory; that is, the theoretical elements have to be separated into logically distinct components. Second, the relationships between these components have to be specified. Bleda and Shackley propose three components: a model of *the cultural construction of risk* based on the theory of archetypes identified in Douglas and Wildawsky; a *model of risk amplification* based upon the general theory of risk amplification in CT; a model of *trust in politically legitimated knowledge* based upon Wynne's (1994) theory of

---

[22] Douglas is very explicit that the origin of her grid/group dichotomy is to be found in Basil Bernstein's distinction between 'restricted' and 'elaborated' codes of speech (Bernstein 1971). We will not discuss the purely sociological merits either of the grid/group analytic or the distinction between elaborated and restricted codes here.

technical alienation. The relationships between these components (or sub-models) are specified as causal.

2. The identified components will have to be decomposed into parameters which interrelate the so derived quantifiable variables. In this instance, formalisation is quantification. The parameterised theory is the model.

3. A formalisable description of the specific values or beliefs attributed to instances of risk will have to be provided. This will be the weighting of the parameters.

4. A formalisable description of the dynamics of stability and change in perceptions and understandings will have to be provided. This description will account the changes in the distribution of perceptions and understandings as well as the salience of differences between perceptions. These dynamics are represented in the flowchart representation which Bleda and Shackley provide (see Figure 4 below, which we will discuss in detail later) in which the causal ordering among the components is set out.

In other words, to build a formal model of CT, a informal, functional and loosely descriptive narrative is re-constructed as determinate, causal formulae. In the rest of this discussion, we will examine how this reconstruction is achieved.

The purpose of the reconstruction is, as Bleda and Shackley say in the title of their paper, to demonstrate how simulation can be used to develop social theory.

> *......operationalising/formalizing some of the most relevant theoretical concepts that currently exist in the literature of risk perceptions. The model provides a conceptualisation and quantitative operationalisation of the dynamics of risk perception underpinned by a solid theoretical framework based on well-established theories and a user-friendly analytical tool that can be employed as a template to run simulations for many different theoretical scenarios. (Op.Cit.* para 2.4)

Bleda and Shackley add a fourth task to the three identified above. They develop a definition of what will count as 'success' for their operationalisation. This definition has two components. The model will be successful if it runs properly, where 'properly' means in expected ways and survives, if not a test to destruction, then at least a strenuous sensitivity analysis. Second, although no body of empirical materials has been gathered by CT researchers for the BSE crisis, there is a body of materials which can be used as a proxy for such a validation test. The model should provide reasonably plausible accounts (in the sense used by Breen cited above) of the materials against which it is calibrated. A successful formalisation of CT, therefore, will not fall over, crash or run amok and will retrodict results which we can plausibly believe are in line with what actually happened during the BSE crisis.[23] What Bleda and Shackley are not claiming (and this is important both for understanding the formalisation and for evaluating their model both as a version of CT and as an account

---

[23] We note in passing, that 'what actually happened' is just one of the things which CT sets out to problematise.

of the BSE panic) is that the model is a description of what actually happened in the BSE crisis. The model aims for 'internal' but not 'external' validity.

It is also hoped that a formal model of CT will be an improvement on the existing theory because it would

> *...(permit) two key theory developments: rigorous testing, refinement, and extension of existing theories that (given their complexity) have proven difficult to formulate and evaluate using standard statistical and mathematical tools; and a deeper understanding of fundamental mechanisms that underpin the dynamics of the phenomenon that the theory is attempting to explain.(Op Cit. para 2.2)*

However, as we have discussed, they do recognise that in accomplishing this they will have to make

> *....a number of explicit choices about levels of analysis, dynamic aspects to focus on, and the classification of relevant parameters. By variation of these parameters within different scenarios it is possible to investigate if the main elements of the theory (which is sometimes under-expressed in formal terms) can be reproduced...(Op. Cit. para 2.2)*

7.4 Varieties of Fomalisation

What exactly are Bleda and Shackley trying to achieve? What is the task of formalisation? We begin with the distinction between formal argument and formal representation. Formal arguments are regimented within a deductive structure. Propositions are derived from premises by the use of formal transformation rules with further propositions derived in turn using the same set of rules. Formal representation is the translation of propositions into a conventionally defined symbolic terms. One formalisation of the proposition 'It is sunny and either we will go for a walk or we will mow the grass' is p v (q ^ r). Given this structure, the structure (p v q) ^ (p v r) is implied. The symbols p, q, r are formal in that they have no reference or meaning of their own. Making the distinction between formal arguments and formal representations allows us to propose a spectrum of formalisation defined by the relative formality of the symbolic language or system used. We will speak of different formalisations points on the spectrum from non-formal to formal symbolisation as quasi-formal and semi-formal representations.

Our second distinction is between formalisation by symbolisation (as above) and formalisation by quantification. The purpose of formalisation by quantification is to bring the deductive machinery of (applied) mathematics to bear on formally defined propositions. Mathematics has rules about the valid manipulation of numerical values. The use of numerals, though, is not by itself quantification any more than the labelling of football positions by allocating numbers to them is. It makes no sense to say that adding the two centre backs together (5 and 4) gives us the central front player (9). The distinction between the two types of formalisation is important because it points to different desiderata for formal representations. Formalisation by quantification aims to achieve preservation of meaning rather than preservation of truth. The meaning of the quantified proposition should be isomorphic with the meaning of its non-quantified version. As we will see, this distinction and the desiderata associated with it are critically important when non-formal qualitative theories are formalised into quantitative ones.

Our third distinction is different again. This is based in George Polya's (1954) observation that although mathematics is concerned with demonstrable reasoning about axiomatised propositions, the propositions it considers do not start out in an axiomatised form. They start out as little more than guesses or 'conjectures' and it is only through the deployment of considerable skill and the use of *plausible* reasoning that they are turned into axiomatised propositions which can then be subjected to the apparatus of formalised mathematical deduction. The various strategies of plausible reasoning which Polya identified (induction, generalisation, simplification, analogy etc) all provide bridges between the original discursively informal guesswork and the resulting demonstrative reasoning. Forms of plausible reasoning provide a set of resources to bring to bear on the challenge of formalisation. These resources are the practices by which scientists, mathematicians, logicians or even sociologists construct formal accounts.

7.5 Patterns of Plausible Reasoning

The argumentative force of demonstrative reason is that it is *truth preserving*. Providing correct methods have been applied at each step in the reasoning, the derived propositions will be true. With plausible reasoning, the argumentative force is different. Here is Polya's summary.

> *In opposition to demonstrative inference, plausible inference leaves indeterminate a highly relevant point: the "strength" or "weight" of the conclusion. This weight may depend not only on clarified grounds such as those expressed in the premises, but also on unclarified, unexpressed grounds somewhere in the background of the person who draws the conclusion. (Op. Cit. vol II p 115-6)*

A little later, he says the process is like this.

> *A plausible argument has been proposed. Each step of it intends to render a certain conjecture more credible and does so following some accepted pattern. Having followed each step of the whole argument, you are not bound to trust the conjecture to any definite degree (ibid vol II p 140).*

Plausible reasoning is about carrying conviction not "machine-like proof". Whilst there are many patterns of plausible inference in mathematics (and elsewhere), three dominate: induction, generalistion and analogy, All three are often used together.

With induction, the truth of some conjectured general law or theorem is tested by enumerating more and more cases. As these cases are tested and found to support the conjecture, so it becomes more credible. The strategy of induction is, then, verification by the consequent. The consequences of the conjecture are examined and verified, or not. For example, until Andrew Wiles' proof, Fermat's Theorem remained an unproven but highly credible conjecture.

Polya adds a further subtlety to this initial position. With the addition of each individual case, inductive generalisation confers, we might say, more marginal credibility. If several consequences are verified, a proposition becomes significantly more credible. To use a vernacular term, if a highly improbable consequence is verified, the theorem/conjecture can be taken to be almost a racing certainty (but remains as yet unproven).

With analogy, things are very different. Along with generalisation, analogy operates in a space of symmetry. In the vertical dimension (so to speak), the argument moves downwards from the class of cases to a specific case. So, for example, we argue from the properties of regular polygons to triangles. This is specialisation. On the other hand, when we argue from the properties of 3 sided figures with internal angels ≤ 90^o to all figures with arbitary angles, we generalise. If a proposition holds true with the transition from the specific to the general, the equivalence of the cases is established.

Both analogy and generalisationn depend upon establishing symmetry relationships. But, at the start, such symmetry is observer dependent and, in an important sense, contextual. It depends upon the purpose in identifying the relationships. The example Polya uses to illustrate this is the analogy between a triangle and tetrahedron. The first is defined by the minimum number of sides for a polygon in 2 space. The latter by the minimum number of sides in 3 space. The credibility of such comparison depends on the *clarification of the analogy*. To achieve such clarification, Polya suggests the following have to be provided:

- demonstration that the cases are governed by the same fundamental laws or axioms, for example those of arithmetic. (eg 0 and 1 are analogous to one another since a x 1 = a and a + 0 = a)
- demonstration that a 1:1 translation can be made between them. This is isomorphism. (eg all natural numbers and all binary numbers can be translated into each other).
- demonstration that an abridgement or condensation does not cause distortion of structurally required properties. Condensation is scale reduction not simplification by means of dispensing with difficult to manage features.

Formal clarification of an analogy, then, has strong requirements.

One approach is to look to a proposition which implies A, the proposition at issue. Call this proposition B. If A is implied by B and B is true, then we can infer A is true. If B is false, all we can say is that A is less credible. Clearly there are cases when the truth of the analogous case cannot be verified, but only affirmed as more credible. This leads to what Polya calls "shaded analogical inference". If A is analogous to B, and B is more credible, then A is somewhat more credible. (Think of this as additive credibility).

As well as induction, generalisation and analogy, Polya identifies other prominent less formal heuristic patterns:

- Examining a related case which acts as a modified version of the problem being examined. Here the credibility depends on the modification and the extent to which properties essential the proposition being examined remain invariant.
- Confirmation from the robustness of generalisation. That is, ignoring what are rare or low probability cases. The proposition A might not be true in this case but most of the time it is......
- Confirmation from the background of general agreement on cases, rules etc. If something is usually taken as unquestionable, we may rely on it (which is not to say it is true, of course simply that we take it as true. This is an important point).
- Simplification is a standard way of making problems more tractable. But it is important to see that the steps from the simpler to the more complex are in fact steps to the same kind of thing; that B is a more complex case of A and not something else entirely. Once again invariant relationships matter.
- Reliance on familiar or widely used patterns of heuristic assumptions. If such assumptions are widely used, they may be taken as reasonably reliable.

The patterns which Polya identifies are not visible in the reconstructed logic of the argument itself. Practising mathematicians know them as familiar features of the mathematician's ways of working. They are purged from the final proof because they are not necessary to ground the constructed formalisation. They are, if you like, a ladder which the mathematics can throw away. They can be dispensed with because context-free formalisation is what practising mathematics comes to. This is not so in applied disciplines. Whilst formalisation might be viewed as a goal, ambition or preference, what defines them is the description or explanation of their empirical phenomena. The plausibility or credibility of these accounts is governed by disciplinary convention. If a disciplinary account of some phenomenon is to be subjected to formal treatment, the adequacy of that formalisation can only be determined from the trajectory of plausible reasoning by which it has been formalised. By itself, the formal representation is not enough.

Earlier, we mentioned earlier that Bleda and Shackley acknowledge that in forming their model they make important choices. Key among them is the selection of formalisation devices to be used. We now look at what such choices might be.

7.6 Formalisation Devices

Any formal account in, be it a formal argument or a formal model, is composed of 'formal objects' which are deployed in various ways. These objects are formatted statements, structures made up from such statements and rules for manipulating statements and structures. These objects have their reference only within the 'world' constructed by the argument or model. Knowing how to use these objects is knowing how 'to do' formal reasoning or modeling. Formal reasoning in applied disciplines also uses such objects (i.e. formatted statements, structures and rules) but the reference for these objects and thus the adequacy of the account given, points both ways; to the world as described by the discipline and to the 'formal world' constructed through the formalisation. In looking at the formalisation process, our focus is on the methods used to maintain this Janus-like character. How are 'facts' and other 'disciplinary knowledge' drawn upon and deployed to fit the materials into the formal structures being used? In Breda and Shackley's case, how does the real world of people's perceptions of BSE, the mass media and Governments' presentations of the outbreak feature in the formal account? How are these two worlds adjusted during the formalisation process so that the formalisations achieve the objectives of meaning preservation and, in Bleda and Shackley's case, retrodiction?[24]

Formalisation devices are a key resource in the process of formalisation. They provide methods or templates for structuring the formal objects of the account being constructed. The degree to which such devices achieve full formalisation varies; some result in quasi or semi formalisations. Naturally the more stringent the formalisation, the stricter the requirements on what can be deployed within the device. Because the case we are examining is a simulation, we will describe three of the more widely used formalisation devices associated with simulation modelling: flow charts, pseudo-code and software (or 'running code'). To draw out the differences between them and to place them along the spectrum of non to fully formal devices, we will use a toy example, that of a simple calculation task.

7.5.1 The Task

Provide a formal description of how to calculate the full cost and the real interest rate of a loan of £10000 taken over 3 years at 7% interest with repayments made every month. At the end of the three years, the debt should be cleared.

7.5.2 Flow Charts

Flow charts represent the information and calculations required to achieve the task as stages or points in a 'flow of data'. They trace the flows of data. Flowcharts have one enormously attractive advantage. They are extremely easy to sketch. They act as facilitating formal devices. Because they are so easy to construct, they allow trial and error testing of structural formats and ordering. This makes them very handy in the first stages of formalisation.

One flowchart of the loan problem might look like this.

---

[24] We will see later that their test of the robustness of their simulation is how well the outputs from the model fit the pattern of risk perception during the actual BSE crisis.

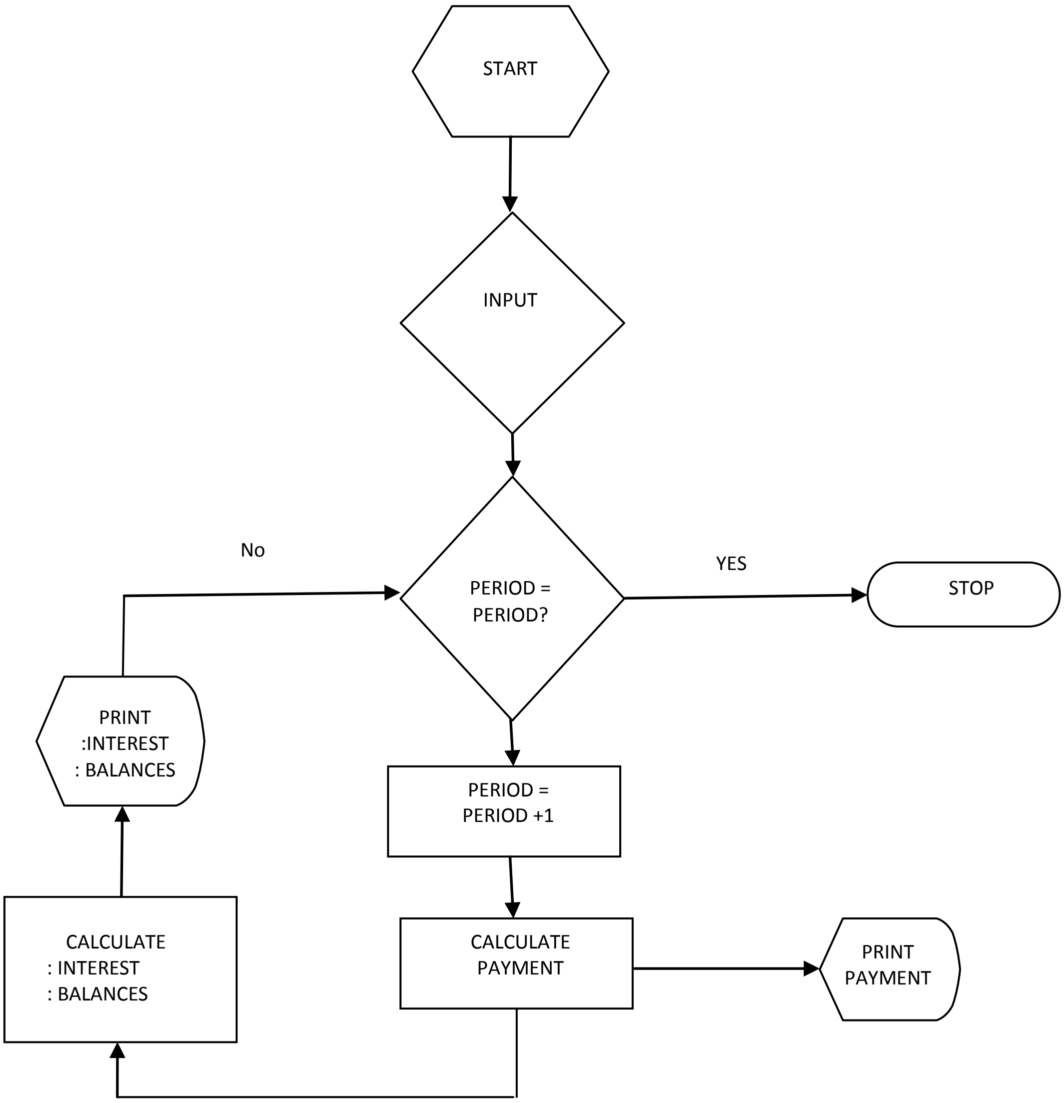


Figure 1 Flow Chart for Loan

The structure represents a flow of data through a set of processing tasks or states. Each of the states is defined by reference to a set of given abstract tasks represented by the various shapes. Thus ellipses are terminal states, rhombuses are data gathering states, rectangles are processing states, and so on. The arrows both connect the states and indicate the order in which they are executed, that is the flow of data from one state to another. The calculations are represented in a single control loop. The loop repeats until the value of the period equals the number of payment periods. Its function is to produce the calculations and statements we need. The print statement/instruction in the loop takes the value of the relevant number from memory and writes it to the screen. This loop is run until the statement which it defines is true. At that point, the whole process stops.

The flow chart does two important things. First, it lifts the representation to an abstract level. The specifics involved in undertaking each step in the task drop out. The task statements are 'context free' i.e.

formal. Second, the syntax and semantics (the grammar) of the graphical notation allows the calculation steps to be compacted into a single iterative loop with a controlling command. Flowchart grammar utilises a spatial top/down left/right reasoning structure expressed in a metaphor of 'flows' of information. The 'gate' for the loop is an evaluative decision point. Flowcharts provide a predefined language comprising defined symbols and the rules for their use. However, the terms in the flow chart are not entirely self-referential. They retain some of the aspects of the task. For this reason, flow charts are semi-formalisation devices.

7.5.3 Pseudocode

Pseudocode is equally semi-formal. It uses a formatted text structure and an idiomatic (or mishmash) style of English. Indentation is used to show the relationships among sub-elements of the task. The pseudocode for the loan example would look something like this.

```
Program To Calculate Loan
  Start
        Set Values
              Set loan balance
              Set loan payment
              Set interest rate
        Variables
              Start balance
              Balance
              Close balance
              Input number of payment periods
        For each period in number of periods do
              Begin
                   Define Start balance
                   Define balance as loan  payment minus Start balance
                   Calculate interest on balance
                   Define Close balance as interest plus  balance
                   Print payment, interest and balances
              End
        Endfor
        Calculate APR
        Print APR
  Stop
```

Pseudocode is the textual analogue of the flowchart (they are logically equivalent) and explicitly gathers all similar functions (defining, inputting, calculating, outputting) in the same place so we can see the structure of the relationships. It also indicates where further decomposition and precision will be required (as for example in managing the savings balance accumulations). The indented structure shows the internal relationships such as the calculation loop bounded by the 'For' and 'Endfor' statements and the calculation steps bounded by 'Begin' and 'End'. The 'For' statement invokes the loop transformation rule as we discussed with regard to the flowchart above. Once again, the 'Print' statement invokes a re-write rule.

The terms used are somewhat but not entirely abstract and their precise meaning is defined as the process is run through. We can figure out roughly what the terms mean from the words that are used. Idiomatic English and indented structure make 'eyeballing' the code a central means of understanding what is happening. Moreover, in more complex problems, often flowcharts will be used to sort out the overall structure (what is doing what to what and when) and pseudocode used to provide a sketch of the coding structure that will be required to realise the flows.

The advantage of flowcharts and pseudocode is that you don't need much (or any) technical expertise to follow them. The boxes and arrows, top-down indented text carry the eye through the structure. Clearly, though, a more detailed understanding is required of the 'language' of each to build a fully formal account.

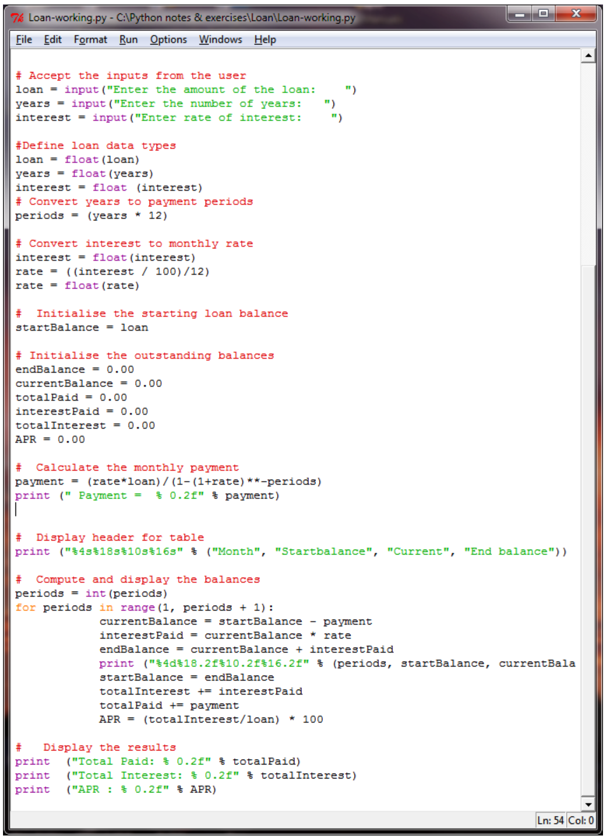
7.5.4 Running Code

Although there are debates about whether any program can be fully proved, there is no doubt that the code in which programs are written is formal. A program is an algorithm or set of connected algorithms designed to solve a general class of problems. To demonstrate how running code acts as a formalisation device, we have built a Python program to solve our calculation problem. Unlike earlier procedural languages, Python is object oriented. That is, rather than being conceived as a set of related procedures or tasks, it conceived as sets of related objects and associated methods. Both, of course, are metaphors. This difference between these metaphors affects the way the code is built and displayed as well as the freedom the programmer has in designing the program. Whilst a Python programmer might have more freedom than a procedural programmer (and how much freedom is hotly debated across these 'tribes') nonetheless Python is very tightly controlled. Only in-built or explicitly and properly defined objects are recognised by the interpreter. For example, in the code set out in Figure 1, the object 'loan' is defined by the number type allocated to it by the programmer. Its value is defined by the user. On the other hand, the object 'input' is defined within Python itself. A lot of programming time is expended on getting these definitions right and in using the methods correctly with the specially defined objects.

Figure 2 is a shell window showing the relevant Python code. The first thing to notice about this code is just how little of it is actually given over to working out the payment and balances. Working out the payment is achieved by a single line of code (although to be fair, this formula would allow us work out the payment readily enough with a reasonable calculator). The meat of the code is the loop labeled "Compute and display balances". However, looking at the loop does not, of itself, tell us how it is going to work..

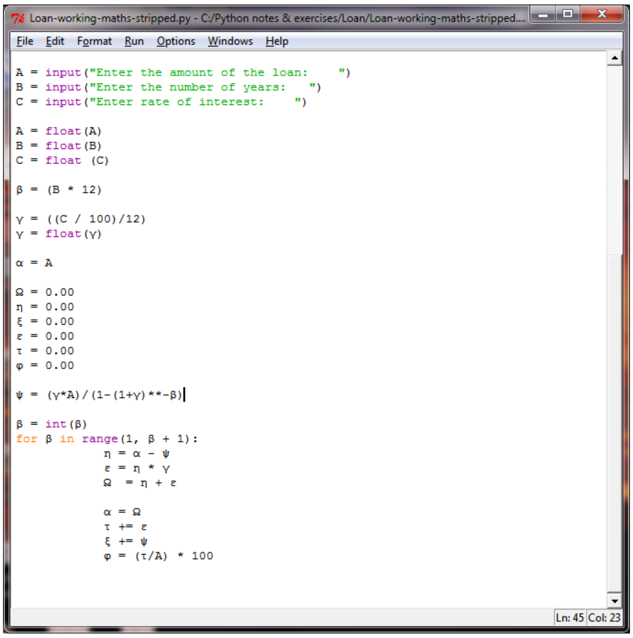
To see what is going on, we have to know that the transformation rule expressed by the FOR loop. This counts from 1 to some defined number by testing whether the current number = the defined number. If the current number ≠ the defined number it executes the code associated with it and increments by 1. The fact it actually counts to the defined number + 1 is an oddity of Python. You have to know that or your code won't work properly and you won't be able to follow exactly what's happening. The code in the loop manipulates pre-defined objects, so to understand what such objects are, we have to refer back to those definitions. The vast majority of the rest of the code, is defining, re-defining or re-writing those objects. Part of this work is necessary because of a decision to keep the user input to integers (and in particular not to have the user input the rate of interest as a monthly decimalised percentage (ie as 0.00). Python is very rigid about how it deals with numbers.

Figure 2



```
# Accept the inputs from the user
loan = input("Enter the amount of the loan:    ")
years = input("Enter the number of years:   ")
interest = input("Enter rate of interest:    ")

#Define loan data types
loan = float(loan)
years = float(years)
interest = float (interest)
# Convert years to payment periods
periods = (years * 12)

# Convert interest to monthly rate
interest = float(interest)
rate = ((interest / 100)/12)
rate = float(rate)

#  Initialise the starting loan balance
startBalance = loan

# Initialise the outstanding balances
endBalance = 0.00
currentBalance = 0.00
totalPaid = 0.00
interestPaid = 0.00
totalInterest = 0.00
APR = 0.00

#  Calculate the monthly payment
payment = (rate*loan)/(1-(1+rate)**-periods)
print (" Payment =  % 0.2f" % payment)


#  Display header for table
print ("%4s%18s%10s%16s" % ("Month", "Startbalance", "Current", "End balance"))

#  Compute and display the balances
periods = int(periods)
for periods in range(1, periods + 1):
            currentBalance = startBalance - payment
            interestPaid = currentBalance * rate
            endBalance = currentBalance + interestPaid
            print ("%4d%18.2f%10.2f%16.2f" % (periods, startBalance, currentBala
            startBalance = endBalance
            totalInterest += interestPaid
            totalPaid += payment
            APR = (totalInterest/loan) * 100

#   Display the results
print  ("Total Paid: % 0.2f" % totalPaid)
print  ("Total Interest: % 0.2f" % totalInterest)
print  ("APR : % 0.2f" % APR)
```

The rest of the definitions simply set up objects as empty slots in which the results of calculation can be written. Whilst the terms used do carry some connotation of financial management, they don't have to. We could have named them after the Royal Family, the England cricket team or a random set of symbols. The terms are entirely internally referring. We used the names we did so that *we* could track what was going on relatively easily and in the hope you would be able to too.



```
A = input("Enter the amount of the loan:    ")
B = input("Enter the number of years:   ")
C = input("Enter rate of interest:    ")

A = float(A)
B = float(B)
C = float (C)

β = (B * 12)

γ = ((C / 100)/12)
γ = float(γ)

α = A

Ω = 0.00
η = 0.00
ξ = 0.00
ε = 0.00
τ = 0.00
φ = 0.00

ψ = (γ*A)/(1-(1+γ)**-β)

β = int(β)
for β in range(1, β + 1):
            η = α - ψ
            ε = η * γ
            Ω  = η + ε

            α = Ω
            τ += ε
            ξ += ψ
            φ = (τ/A) * 100
```

Figure 3

If, as in Figure 3, we change the object names to Greek symbols we wouldgenerate something that looks just like a piece of formal mathematics and which runs quite happily, although it won't print any results!

Python provides a set of pre-defined objects (formats, structures and rules) and in using them, the formalisation extracts the content from the semi-formal or informal representations of the solution procedure and casts it into an abstract 'language'. The work of formalization is using a variety of translation rules to map the representation into the chosen formal language. It is only the relative familiarity of the terminology in the Python code that allows us quickly to eyeball it and see what is going on (we are familiar enough with the referents). When cast into purely abstract terminology and stripped of the handrails of guiding comments, it becomes impossible to follow as a set of reasoning about a loan. However, (providing we can hold all the symbolic definitions in our heads and understand the built-in Python terms!), with enough training and practice it is quite followable as a formal structure. That is because its objects have become entirely self-referencing.

7.5.5 STELLA

Bleda and Shackley use the STELLA modeling platform for their formalisation. STELLA is a simulation language based on a metaphor of reservoirs or pools of stock and rates of flow in and out of them. STELLA uses a flowcharting interface to represent the flows and is particularly suited to domains such as ecology, applied biology and business where phenomena can intuitively be conceived in a stock/flow way (e.g. populations and resources: consumption and predation; cash and goods; costs and revenues). Having specified the relationships with the graphical modeling tool, that is having set out the links between the factors which control the rates of flow, the rates themselves are defined as sets of equations. These equations plus the definitions of the values for the various entities provided either by the theory or case being examined constitute the code base for the simulation. In STELLA, then, we have a layering or hierarchy of formalisation devices. The graphical objects defined within the flowchart based modeling language are re-written by STELLA into the code that runs the simulation itself. The definitions of object type and 'method' associated with each of these objects are subject to a variety of transformation rules to produce the values re-written into the equations. It follows that when working with STELLA, a number of constraints will be in operation. The grammar of the entities (how they can be related) is highly specified. As with Python, we can only build a model the way STELLA insists models be built.

8.0 COWCULT AND THE FORMALISATION OF CT[25]

Bleda and Shackley seek formalisation through quantification. Their model is meaning preserving rather than truth preserving. The propositions of the model preserve the meaning of the CT theory they draw upon. As we saw earlier, the propositions of CT are indefinite, non-formal and functional. In the formal model, these will be analogised as a set of quantified, causally related formulae. To understand how Bleda and Shackley do this,

---

[25] We are very grateful to Mercedes Bleda and Simon Shackley for allowing us to use their model and for providing access to the underlying STELLA model code they constructed.

we will look at one particular example, the specification of archetypes and their dynamics. The empirical validity of the types is a premise for the model. Its aim is not to test their explanatory value, but to use them to construct on explanation of the patterning of risk perception during the BSE crisis. This is not a weakness in the formalisation. The fit of the archetypes to the BSE case is assumed *for the sake of the model*.

8.1 Simplification

Bleda and Shackly reduce the complex and abstracted qualitative argument at the heart of the Douglas and Wildavsky theory of risk to a small number of constructs. The first of these is the definition of the interaction effect between 'context' and 'perceptions' or attitudes. The second is the patterning of those attitudes. In Douglas' theory, the relationship between grid/group context is an open one with the interaction being one of the mutual shaping of perceptions and context, Bleda and Shackley resolve this indeterminacy by dropping one side of the relationship. The types are simply taken to be inter-related clusters of congeries of beliefs, values and attitudes. The role that the grid/group theorisation plays for Douglas and Wildavsky is set aside. In its stead the sets of propositional attitudes are the assumed archetypes. The archetypes are not defined in terms of the CT machinery that produced them nor in terms of the patterns of attitudes, perceptions and beliefs found in the BSE crisis. They have become entirely self-referential. What we mean by this is not that Bleda and Shackly re-invent the archetypes giving them properties that would not be recognised by CT. Far from it! They use CT to characterise each of them. It is just that this characterisation is independent of the grid/group theorisation at the heart of the Douglas and Wildavsky theory; the theory which gives them meaning.

The second element in the reduction is the specification of the relationship across the components which 'influence' perceptions of risk. Douglas and Wildavasky see this relationship as one of the mutual entanglement of reality and culturally shaped perception. Bleda and Shackley re-cast this as a quantified weighting. Their model is built around such weightings (within and between archetypes). The model is, in effect, a complex weighting structure. For this to happen, the subtle functional co-relationships outlined by Douglas and Wildavsky are turned into a straightforward causal consequences.

The reduction is not, then, an abridgement in Polya's sense. There is no 1:1 mapping between the components of the original theory and the elements being used in the formalisation. Instead, we have analogised simplification. The question is whether this simplification retains enough structural symmetry for it to be meaning preserving. Does the pared down version of CT retain the explanatory or descriptive force of the original? If sufficient structural symmetry is retained, the first step in formalisation has been successful. If it has not, then right at the start, 'meaning drift' has set in. With meaning drift, the invoking of the vocabulary of the types would not fully preserve their meaning and a bifurcation of meaning between archetypes as descriptors of empirical phenomena and archetypes as formal objects will have begun.

8.2 Causalisation

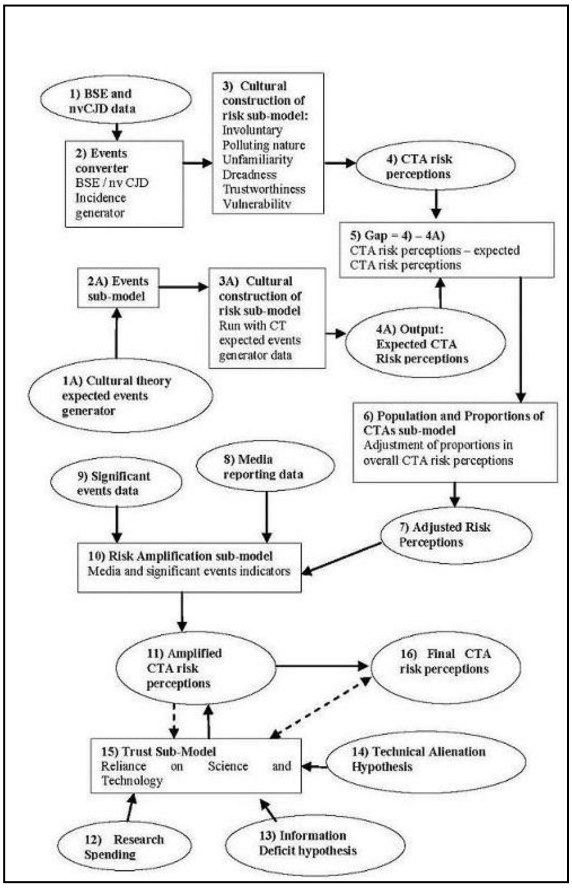
Prediction in CT in general, and especially in the Douglas and Wildavsky version, is deliberately under-emphasised. The relationships of social structure and cultural phenomena are co-variations. When grid/group relationships are of *this* kind, we can expect *those* kinds of beliefs, values and attitude; and *vice versa*. Bleda and Shackley want their simulation to retrodict the pattern of risk that occurred during the BSE crisis. To become a predictive theory, CT has to be transformed from being descriptive of mutual elaborations to being descriptive of causal sequencing. Once CT is couched as causal sequencing, structuring the causal relationships allows prediction. The STELLA platform, built as it is around the metaphor of stocks and flows and the causal processes affecting the rates of flow in and out of particular stocks, *requires* causalisation. It also provides the machinery for developing predictions once causalisation is achieved.

Before predictions can be derived, though, CT has to be re-formulated as a causal theory. Here is Bleda and Shackley's reformulated structure.

1. Related sets of external events generate specific information about particular possible risks.
2. Individuals hold patterns of beliefs, values and attitudes (including beliefs, values and attitudes about risk) which conform to a small number of types. Membership of a specific type determines the attitudes any particular person adopts to possible and actual risks.
3. *Ceteris paribus,* the patterns of beliefs, values and attitudes will determine the scale of any change in the perception the risk consequent upon the acquisition of the information.

The purpose of the simulation is to predict how such movements will occur, demonstrating the variety of attitudinally driven perceptions of 'the same evidence' about BSE. This reformulation involves more than a simple re-writing of the core dynamic of the Douglas and Wildavsky theory. The causal structure has been derived from Douglas and Wildavsky but also transforms it. In making the derivation, the notion of causal force represented by weighting of the parameterised variables is added. The outcome is represented in the cultural construction of risk sub model in the flow diagram set out in Figure 4. This is the semi-formal representation of the formalised model which is to be built.

Figure 4



8.3 Operationalisation

Conventionally operationalisation involves three steps:

1. Terms or constructs in a theory are re-defined as determinable parameters. The parameters are proxies for the constructs;
2. The parameters are specified in terms of empirically identifiable and measurable variables;
3. The variables are defined in terms of sets of measures of various types.

The task of operationalisation begins with *parameterisation.* Once that is achieved, the theory can be transformed into quantifiable statements using variables. The flowchart is an organisation of the components in the CT model. It is not a specification of them as an operationalised causal flow. In the flowchart, CT objects are labels arranged in the usual top/down, left/right sequential order. With operationsalisation, Bleda and Shackley define what the labels in the flowchart stand for. This is achieved by carrying out two steps at once. The constructs are re-written as parameters, variables and measures and these new theoretical objects are arranged according to the logic of the STELLA modeling structure. The labels in the flowchart are transformed into concatenations of STELLA-defined modeling objects; that is, combinations of flows and stocks with connectors and converters acting upon those flows and stocks. These concatenated objects are the parameters of the causal CT whose meaning is defined in terms of the variables which realise them and whose measures provide the quantification necessary. The theory's statements expressed in the causal structure of the flowchart are transformed into derived quantitative statements built from the STELLA objects.

To bring out how this happens, we will trace through the first steps in the causal structure; the operationalisation of 'BSE events' and 'CJD events' as perceived by the various types. Here is Bleda and Shackley's definition of the task and its solution.

> *Since the interpretation that each of the four CTAs makes of the accumulation of BSE and nvCJD events over time is different, their influence on the construction of risk perceptions will also differ for each CTA. In order to account for this effect, the model assigns different weights to the same number of accumulated events for each CTA.*

Earlier, we summarised the first two steps in the causal structure as:

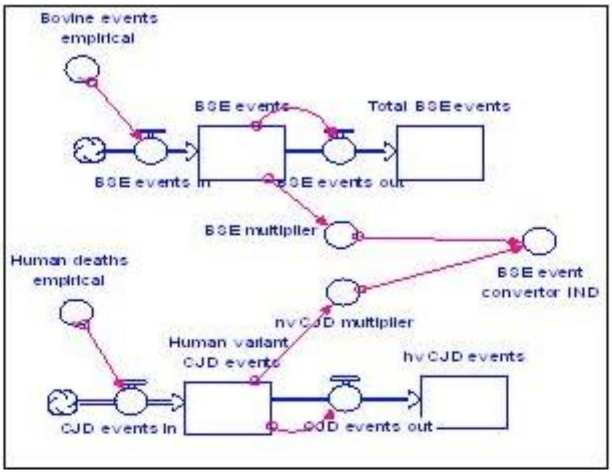
1. Related sets of external events generate specific information about particular possible risks.
2. Individuals hold patterns of beliefs, values and attitudes including beliefs, values and attitudes about risk which determine the attitudes they adopt to possible and actual risks.

Using STELLA modeling techniques, these two are converted to quantified propositions about the accumulated numbers of relevant BSE events perceived by the different types. The flow of events goes like this.

1. Individual events related to BSE and human deaths from nvCJD occur in the external world on a continuing basis.
2. The rates of occurrence of these events are the inflows of each type of event.
3. At any point there is a running stock of nvCJD and BSE events and an accumulated total of both sets of events. These are 'the facts' about nvCJD and BSE.
4. Although 'in reality' there is a single rate of flow for each of the BSE and nvCJD events, the different archetypes perceive the running stock (ie those events which are current) differently. (Bleda and Shackley take this to be an axiom of CT). Each current stock is, therefore, subjected to an archetype-specific modifier (the 'multiplier' in the model). The integration of the values for each type becomes the BSE converter for each type. That is, they become the perception of the facts for each type.

The semi-formal representation of the flow chart is reconstructed in the STELLA modeling language first as a general representation of the flow of events relating the BSE and nvCJD as set out in Figure 5 and then as perceived specifically by the individualist archetype set out in Figure 6.

Figure 5



The full formalisation of the individualist event converter is this

```
        BSE_event__convertor_IND =
IF(nvCJD_multiplier=0)THEN(BSE_downscaler)ELSE((BSE_downscaler*0.001)+(nvCJD_multiplier*0.
999))
```

This is the quantified formalisation of the individualist archetypes perception of the number of BSE and nvCJD events. The weighting is the combination of a 'downscaler' for BSE events and a 'multiplier' for nvCJD. In the individualist case, the relevant quantities are 0.01 and 0.999.

Of course, no quantities or measures have been fed through from CT to the causal reformulation. Consequently, quantification is achieved as part of the operationalisation rather than *vice versa*. The accumulating current stock of events is defined as those in the public record.[26] The modifiers or multipliers allocate weightings according to their assumed relative values *vis a vis* each other. Each type is characterised as holding a uniformly valued attitude towards these BSE and nvCJD events. The egalitarian type, for instance, is allocated the highest multiplier for BSE because *it is assumed* to care most about animals. On the other hand, the individualist type has the highest multiplier for nvCJD deaths because *it is assumed* to care most about human life. The other types are placed between these two. The resultant values are not precise measures. Rather, as Bleda and Shackley admit, they express a subjective ordinal scaling somewhat equivalent to 'a little', 'some', 'more', 'a lot'. The BSE event converters which are produced by this operationalisation are indeed expressed as the numbers of BSE events perceived by each type. But, since these are ordinal measures, it is not exactly clear what applying any mathematical operation on them would actually mean. On the other hand, it is clear they the formulation does preserve a great deal of the original CT conception of what an archetype means. Meaning drift may have started but it has not gone very far.

The next step in the model is to turn these 'perceived events' into 'perceived risks'. To repeat, there are no empirical data on the clustering of attitudes towards the BSE panic to shape the way that this transformation might go. To provide the bridge needed, Bleda and Shackley analogise the case of BSE to other

---

[26] We say nothing here about the processes of 'public definition' of events as nvCJD or BSE though undoubtedly most versions of CT would wish to.

cases discussed in the psychometric and sociological literature. The exact degree of symmetry between this case and those cases is, however, left unaddressed. As a consequence, in Polya's terms, the plausibility of the reasoning must be reduced. Without some measure or determination of symmetry, we cannot tell whether we have similarity, congruence or isomorphism.[27] Nonetheless, from this literature, Bleda and Shackley extract a number of 'dimensions' (that is parameters) along which the different types will differ in their attitudes to risk. The parameter 'attitudes to risk' is defined in terms of these variables. For each type, the level of risk of an event is fixed by the following; involunariness, polluting nature, unfamiliarity, dreadness, trustworthiness of Government, vulnerability and fairness. Using thumbnail sketches as the rationales offered for each type, and following the same logic as used with the event multipliers, each is allocated a weighting (a position on an ordinal scale) on the basis of an interpretation of the location of their assumed attitudinal bundle along that 'dimension'. These weightings are the values used to turn the BSE events generated earlier into stocks of perceptions of risk. Thus a quasi-formal representation of BSE events is modified by a non-formal representation of risk. The outcome, therefore, cannot increase in its formality.

The net result has the appearance of a kind of 'attitudinal soduku'. There is a distribution of values on each dimension for each type. For each dimension, the total of these values sums to 100%. However, each of the dimensions is allocated a comparative weighting within the set of dimensions. This 'comparator weighting' is based on the values allocated to it for each type. So, while the fairness and dreadness 'score' are highest for egalitarians, for individuals, the highest 'score' is fairness. The combination of dimension scores gives us a profile for each type. But there is also a profile across the 'dimensions' themselves with fairness emerging as the most highly weighted and unfamiliarity the least. This profile is distinctly bi-modal, with fairness and trustworthiness of Government being much stronger than the other 5 dimensions. The empirical validity for this distribution is remains undetermined. Here is the rationale offered for the individualist and fatalist profiles.

> *Individualists rely on their own sense of trust in others, hence weight for trustworthiness is high. Individualists believe equality of opportunity is vital, hence fairness is also highly valued.......*
>
> *Fatalists will be most concerned about fairness - they think that they are always on the receiving end of risks. They also regard themselves as vulnerable to risks being imposed on them. Of less importance to fatalists are involuntariness (rarely able to make it otherwise), trust (low in any case) and polluting nature.*

The resulting distributions of weightings are as follows:

---

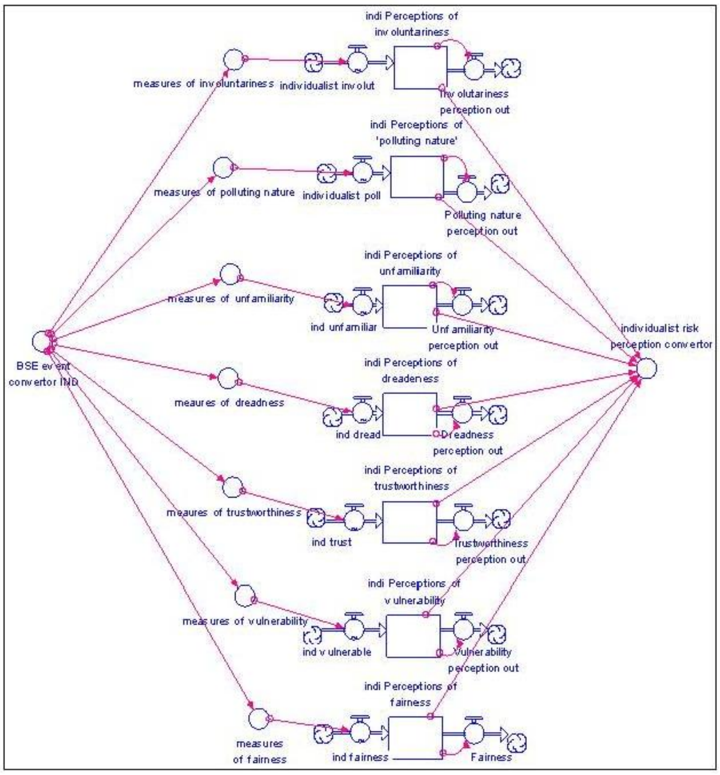
[27] What is interesting here is that (inadvertently?) Bleda and Shackley treat the BSE panic an inductive evidence for the validity of the variables they extract from the literature. Using Polya's terminology, this then becomes a buttress for the credibility of the use of the variables in the BSE case.

Dimension	Individualist	Fatalist
Involuntariness	0.5	0.5
Polluting Nature	1	0.2
Unfamiliarity	0.6	0.7
Dreadness	0.4	1.7
Trustworthiness of Government	2.15	0.3
Vulnerability	0.2	1.8
Fairness	2.15	1.8

The weighted profile attributed to each dimension profile is, of course, important because of the implications it has for the way the simulation actually works.

The modeled perception converter for individualists is given below.

Figure 6



Once we have understood how the parameterisation has been undertaken, we can see that the model assumes that the risk evaluation of BSE event events by the individualist type will overwhelmingly be in terms of the trustworthiness of government and fairness. These comprise 70% of the weightings for that type.

The formalisation of the individualist risk converter is

```
individualist_risk_perception_convertor =
((indi_Perceptions_of_’polluting_nature’+indi_Perceptions_of__dread
eness+indi_Perceptions_of__fairness+indi_Perceptions_of__involuntar
iness+indi_Perceptions_of__trustworthiness+indi_Perceptions_of__unf
amiliarity+indi_Perceptions_of__vulnerability)/7)/10
```

Applying this converter to the modified value of BSE events for individualists provides the modeled individualist perception of risk. The steps are replicated for each of the other archetypes. Once this process is complete we have *a number* that stands for the perceptions of the risk attached to BSE events for each type but just how formal this representation is of the archetype's risk perception remains unclear.

Mutatis mutandis, the same approach is used for each of the other sub-components of the model. Each component is decomposed into causally structured parameters with associated variables. For each variable, numerical values are derived through a process of subjective rationalisation. The causal force of each element in the model (and hence the whole operation of the model) is a direct consequence (as we have just seen with the perception of risk for the archetypes) of these rationalisations.

8.4 Summary

Our purpose in this discussion has been to trace through the process by which a loosely formulated and discursive theory is transformed into a formal one. The formal theory is a direct analogy of the discursive one. In contrast to its use in formal logic, formalisation is not translation into abstract symbols but into quantified values. Quantification is the formalisation process. Whereas successful formalisation in logic and mathematics is truth preserving, we have suggested quantified formalisation in the applied sciences is meaning preserving. The re-writing and derivations that are required by formalisation should preserve the meaning of the original theory. This gives us two criteria for the adequacy or plausibility of the reasoning being undertaken; the adequacy of reference for the quantified variables and the credibility of meaning preservation across the transformations. Working through the COWCULT model (or, at least, a part of it) we have shown how by a process of reasoning though analogy and the application of re-write and derivational transformation rules, Bleda and Shackley construct a model of a core element of CT. This model produces numerical statements of the profile of risk perceptions for different types of individuals. However, we cannot say that these numerical statements are fully quantified and so no fully formal. They have more in common with the semi-formalisms of flow charts and pseudocode. The advantage of not being fully formal is that the representation largely preserves the meaning of the original qualitative CT theory. The new theory may not be formal but significant meaning bifurcation has not occurred.

As we have worked through the reasoning, we have noted a number of critical junctures; the simplification of the original Douglas grid/group CT; the causalisation of the simplified theory; and the achievement of formality through operationalisation. At each of these junctures, we can see the double-fitting or mutual accommodation which was required so that the successively morphed CT theory could be fitted to the formalising apparatus of the STELLA model. This double-fitting is the essential work of formalisation.

9.0 COWCULT, CT AND FORMALISATION

The COWCULT model takes sets of values (counts) of BSE infection among cows, nvCJD diagnosis among humans, numbers of articles in the media regarding both as well as patterns of spending across Government departments and, using differentiated weightings for the set of cultural archetypes assumed to characterise the population, produces measures of the relative predisposition of the archetypes, first, to see events as BSE ones, second to see the risks related to BSE events as increasing, decreasing or stable and third to show how trust of Government pronouncements about the course of events underway changes during the outbreak. Running the model for time series data of the relevant counts produces a changing pattern of risk perception for each archetype and an overall level of risk perception. The aim of running the model is to facilitate the tuning of the weightings in the light of the generated data so that the simulated patterns of risk perception can more closely resemble the patterns of risk which were observed during the crisis. If this tuning were to be successful, the resulting tuned model could offer insights into how CT might be re-shaped as a formal, predictive and generalisable one.

Bleda and Shackley set themselves a number of objectives against which COWCULT might be assessed. Broadly, these fall into three categories:

1. The degree of formalisation achieved;
2. The internal validity of the model;
3. The contribution to the development of CT

External validity (i.e. how well the model approximates as a description of the actual forces and processes to be observed during the BSE crisis) is not an objective. There have been no studies, and certainly no studies providing the required detail, to allow this assessment to be made. As such, the COWCULT model cannot be viewed as an *idealisation* of the BSE crisis in the way that the Gas Laws could be said to provide an idealised model of the behaviour of bodies of gas. We have no measures of calibration for the model. Inevitably, as we will see, this must weaken any implications the model has for CT.

9.1 Degree of Formalisation

COWCULT seeks formalisation through quantification and the model certainly takes quantified data of events, reports and spending as its inputs which are then modified by having weightings for the archetypes applied to them. The question is just what the resulting quantities actually mean. That question turns on what the weightings mean. As we saw in the two examples we examined, because we have no input data for the parameterised variables specified for the archetypes, the numbers are, in fact, basic ordinal measures. They simply express what Bleda and Shackley *think* is the order that that the archetypes ought to stand. Given that there are seven 'dimensions' specified for the attitudes of the archetypes, Bleda and Shackly might just have easily ranked them high, high medium, medium high, medium, medium low, low medium, low and attributed a scale of 1 - 7 to them. It is the apparent precision of the quantification that provides for quantified formalisation. Bleda and Shackley's quantification cannot claim to be precise. There is quantification but it is only minimally formal.

The second criterion of formalisation is meaning preservation. Here the question is somewhat different. Since CT is entirely non-formal and discursive, Bleda and Shackley have to define what they take CT to be in order to decompose its elements and formalise their description. This is the specification of the component sub-models and the parameters for the archetypes. Treatment of these specified elements does preserve meaning across the model. The coded statements do preserve the meaning of the original specification. However, the potential disjuncture is between CT as a species of sociology theory and CT as specified by Bleda and Shackley as input for their simulation. If meaning is not preserved across that transformation then the resulting model cannot be a fully formal model of CT as a theory in sociology. We return to this question below.

The third criterion is the extent to which the propositions of the formalisation are context free. Are the meanings of the symbols used in the formulae independent of the context in which they are deployed? Here, as a cursory review of the code statements we have cited above will show, we can be unequivocal. The terms of the formulae depend for their meaning on the context within which the model is being deployed. In other words, without considerable re-building, the COWCULT model could not be applied to any other similar 'panic' situation let alone any other long running national event. Modelling public attitudes towards the Olympics (say) would require a completely new model, with new terms and formulae not just new weightings. We suggested in our discussion of the STELLA code included in the model that it was more like pseudocode than fully formal description.

9.2 Internal Validity

Internal validity is assessed against two separate criteria; the internal consistency of the logic of the model and the model's stability. Bleda and Shackley present results on both counts. The model was tuned over several iterations (as one might expect) and executed for data that covered a 16 year period. The resulting patterns of risk perception and amplification were entirely explicable in terms of the assumptions which Bleda and Shackley made. To put it in the terms we used earlier, the model did not crash or run amok. Bleda and Shackley run a number of alternative scenarios to test the sensitivity of their modeled variables to variation in input. Once again the output data was entirely explicable within the assumptions of the model. The model seems to be relatively stable. In sum, COWCULT is pretty robust.

9.3 Contribution to CT

While, as we have just seen, the building of COWCULT was demonstrably a technical accomplishment, the purpose of the endeavour was not simply an exercise in simulation. The hope was that the model would offer ways of sharpening CT. There are two issues to be considered in this regard. The first is the extent to which gathering the kind of data required to develop COWCULT as an empirically valid description of any CT phenomenon such as the BSE crisis is likely to be feasible in the near future. The second is the extent to which CT as defined in COWCULT maps onto CT as deployed as a sociology theory.

The first issue requires the empirical validation of the archetype structure. In the Douglas and Wildawsky formulation, the archetypes are derived from the cross comparisons of strong and weak influences of the grid/group structure. They are, then, logical constructs, not empirical ones. For COWCULT to make a contribution to CT, the archetypes themselves have to be substantiated. This will require studies of a type and on a scale not, as yet, undertaken within CT. In turn, to mount these, significant development of the supposed characteristics of 'grid' and 'group' ties will have to be achieved. In short, the rationalisation of the supposed interplay of grid and group influences will have to become a substantive theory.[28] If the archetypes are validated, it will be possible to derive empirically grounded parameters and measurable variables for them and to design studies (surveys, experiments, ethnographies or whatever) to characterise them precisely. Without all this work being completed *within CT*, COWCULT cannot make any meaningful contribution to CT theory.

The second consideration requires a judgment to be made with regard to the meaning preservation of the transformation of CT from an informal, discursive functional narrative into a formal, causal structure. This is brought sharply into relief by considering the fulcrum around which each account turns. For CT the relationship between perception and social context in sociology is mediated through the contextual co-variation of classification schema. Neither is prior; both are mutually explicating. We used the Weberian term "elective affinity" to describe this reciprocity. COWCULT 's CT sees the contribution of perception and social context as causally weighted. The weightings are given in the variables defined for the various components of the archetypes. This is an entirely different conception, with the weighting structure providing a mechanism which produces the pattern of risk perception generated by the model. Elective affinities and causal mechanisms are entirely different conceptions of the relationship. In that sense, turning CT into a causal model transforms its central conception replacing it with one which is amenable to modeling in STELLA. COWCULT CT is not sociology CT.[29]

9. 4Conclusion

Clearly COWCULT is a technical achievement. The model runs in a stable manner and the outputs are coherent in terms of the assumptions made. However, the mapping of CT in COWCULT to CT as practised within sociology remains not so much loose as underdetermined. The key problems to be resolved lie with the work that

---

[28] Whether this would be a 'theory of the middle range' of the kind AS is seeking, we cannot say.

[29] Though it *might* be AS CT.

researchers within CT will have to complete to re-construct CT as a predictive causal theory and the detailed studies that will need to be carried out to characterise the components of that causal theory in ways that are empirically grounded. Only when both of these steps have been taken will it be possible to see if the COWCULT model can actually contribute to the further refining and sharpening of the re-constructed CT.

10.0 AGENT-BASED MODELS AS SOCIOLOGICAL ANALYSIS

COWCULT is not strictly an agent-based model. It does not model the actions of individual actors. We chose it as our example because it is a clear and systematic attempt to struggle with the problems of formalising an informal sociological theory. In the end, our conclusion was that although COWCULT is robust as a simulation, as a form of CT it left a lot to be desired. To achieve the translation to formality key elements of the Douglas and Wildawsky theory had to be amended, re-constructed or simply ignored. As a result, the mapping between the two forms of the theory could not be said to be symmetric in the sense required and the simulation, is not, therefore, a strict analogy of the Douglas and Wildawsky theory. In our view, unless the Douglas and Wildawsky theory can be re-constituted sociologically so that it has a form more amenable to formalisation, this is highly likely always to be the case. So, while COWCULT is a good illustration of the challenges in formalising sociological theory, it is not a good test of the sociological value of agent-based models themselves. In contrast, we will now examine a model which has been constructed to throw light on the issue of supervenience and the bottom up emergence of the macro from the micro, the two key issues in AS. The model uses extant sociological theory as a departure point but does not seek to translate it directly into terms amenable for an ABM. The theory simply defines a model of a 'possible world' to be modeled. In this case, then, our interest is not, as it was with Bleda and Shackley, in how well it translates and preserves the meaning of a sociological theory, but how plausible and insightful its sociological analysis of the scenario it constructs might be. The example we have chosen is the ABM used to account for emergent social communication introduced in Salgado and Gilbert (2012) and detailed in Marchione, Salgado and Gilbert (2010).

10.1 The emergence of social order and the 'double contingency'

In *The Social System*, Talcott Parsons (1951) defined the central topic of sociology as "the problem of order". This problem is one of achieving the co-ordination of expectations when two individuals interact. Unless there is a common definition and interpretation of the situation coupled with conformity to the complementary expectations each has of the other, sustained and successful interaction will be impossible.

> *Expectations, then, in combination with the "double contingency" of the process of interaction as it has been called, create a crucially imperative problem of order. Two aspects of this problem of order may in turn be distinguished, order in the symbolic systems which make communication possible, and order in the mutuality of motivational orientation to the normative aspect of expectations, the "Hobbesian" problem of order. (Op. Cit. p 36)*

Parsons' problem and his framing of the solution have remained central and controversial in sociology ever since. Recently, Niklas Luhmann (1996) has argued that in the absence of shared media of communication, the resolution of the double contingency would be highly improbable. As a consequence, such media must have evolved as part and parcel of the emergence of social life. In an echo of Parsons' analysis, Luhmann defines three sources of this improbability: the co-ordination of meaning; the extension co-ordination to others over space and time; and normative adherence, that is that both individuals agree to conform to the expectations of the other. The media we have evolved to overcome these improbabilities are cultural signs, dissemination media and symbolically generalised communication media. Luhmann, thus, defines a possible world in which asocial individuals faced with the improbabilities of co-ordinated social action evolve media of communication to facilitate that action. From such co-ordinated action, social structures and society emerge. Marchione et al take this possible world and, using ABM, model how this might happen.

10.2 Simulating the emergence of a shared lexicon

The first step that Marchione et al take is to simplify the problem by dropping the third improbability. This does not mean that normative adherence is thereby eliminated from the model. Rather, it becomes a tacit assumption and, as we will see, provides the glue that holds it together. The simulation is now focused solely on the improbability of co-ordinated meaning and the improbability of communication with others over extended space and time. The test which Marchione et al set themselves is to simulate the emergence of a shared set of cultural signs (a 'lexicon') within a population through a process of extended communicational 'reach' beyond the immediate dyadic interaction. They see the former as a proxy for the emergence of speech and language whilst the latter a proxy for dissemination media such as writing, broadcasting and so on.

To frame the simulation, Marchione et al set out how a term/sign becomes part of the lexicon and how it becomes widely shared. The first is a consequence of the frequency with which a particular referent is used in interactions (that is how often 'the topic' is communicated about). Some individuals will communicate about a narrow range of matters, thus changing topics infrequently while others will change topic frequently. In like manner, some individuals will be connected to many nearby actors and able to communicate with them, and others will not. The extent of this connection is defined as the number of network connections an actor has. Some will have many, others will not. Those with many will be said to speak loudly whilst those with few will be said to speak softly. This dimensional pairing (frequency of topic change and density of network) gives Marchione et al a set of "paradigmatic communication strategies" to model.

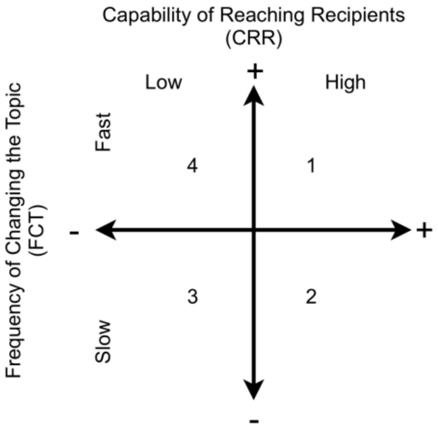


Figure 7 Paradigmatic Strategies

Luhmann's possible world is operationalised by definining the following ontology; actors and objects; communicative interactions; and the lexicon. Actors are speaker/hearers and objects are putative topics of communication. Communicative interactions consist of the exchange of 'words' about a set of defined objects scattered in the world (flower, tree, leaf, etc). Actors belong to 4 groups of 10. They can see objects and move in the world. Utterances are stipulated to be about one and only one object which can be seen by the speaker and hearer(s). Utterances are given a loudness measure in terms of the radius within which they can be 'heard' (i.e. the number of hearers who can also see the object). The larger the audibility radius the 'louder' the utterance. An actor's movement is set by the number of steps it takes at each turn of the simulation. The more frequent the steps, the faster change in topic is likely. Finally, objects are placed at regular intervals so that the shorter the step the longer it is assumed agents speak or hear about it. Communication is only about the object that can jointly be seen (ie both must have a free path to the object and the hearer must be within the speakers audibility range). The lexicon consists of words (w) of 4 letters with a marker for the origin of the word and made up from a randomly attributed set of letters drawn from a common alphabet. At each turn of the simulation, words are uttered and heard. The value of a word is increased or decreased depending on the outcome of the interaction. This exchanges are modeled on the "observation game" of Vogt and Coumans (2003) and the whole simulation written in NetLogo. In Vogt and Coumans' "observation game", the process for enlarging both the size of the lexicon and the spread of its use is organised as follows.

1. *Two agents are randomly selected from the population. Arbitrarily, one agent is assigned the role of speaker and the other becomes hearer.*
2. *The speaker selects randomly one meaning as the topic from the ontology and informs the hearer what the topic is, thus establishing joint attention.*
3. *The speaker searches its lexicon for words that are associated with this topic and selects the association that has the highest score σ . If the speaker fails to find a matching association, it invents a new word and adds the word-meaning association to the lexicon with an initial association score of σ =0.01. The word is communicated to the hearer.*

4. *The hearer searches its own lexicon for an association for which the word matches the received word and of which the meaning corresponds to the topic.*
5. *If the hearer succeeds in finding a proper association, the observational game is a success. Otherwise it fails. Both agents know the outcome.*
6. *Depending on the outcome, the lexicon is adapted as follows:*
   - a. *If the game is a failure, the hearer adopts the word and adds the word-meaning association to its lexicon with an initial association score of $\sigma = 0.01$. The speaker lowers the used association score by $\sigma = \eta \cdot \sigma$, where $\eta = 0.9$ is a constant learning parameter.*
   - b. *If the game is a success, both robots increase the association score of the used association by $\sigma = \eta \cdot \sigma + 1 - \eta$ and they laterally inhibit all competing associations by $\sigma = \eta \cdot \sigma$. (An association is competing when either its meaning is the same as the topic, but its word differs from the uttered word, or when the word is the same, but not its meaning.) (Op. Cit. para 3.4)*

Although Marchione et al change the detail of the interactions in their "chatty game" (in particular, the randomised invention of new words) the outcomes are the same. When an actor emits an utterance, the word uttered enters the lexicon with its frequency value increased or decreased depending whether the word is already in the hearer(s) lexicon.

The measures of the game are defined as follows:

- The hearing capability of a group is the sum of the interactions among members of a group and the other groups. The speaking capability is similarly defined.
- The mutual relation between actors is the mean weighted difference between interactions from agents a to b and b to a.
- Group coherence is maintained by preventing any member moving more than 100 pixel units away from any other member of the group which in turn sets the maximum mutual relation value for the group.
- Finally the probability of any one group spreading its w in the lexicon is a ratio of all the groups spreading their words. the simulation is run (or re-run) from a set of initial conditions with these derived measures.

Over the run of the simulation, the lexicon becomes dominated by words 'invented' by those actors who don't change topic very much and who have the largest communication networks.

10.3 The sociological significance of "the chatty game".

The problem which Marchione et al are seeking to illuminate is the resolution of the 'double contingency'; the achievement of shared definition of the situation and normative compliance in the absence of cultural and social systems. This is Parsons' problem of order. Following Luhmann, they treat it as a problem of coordinated communication. How does this get started and stabilised into an emergent (i.e. supervenient) social structure? Although Marchione et al want to do this without getting embroiled in philosophical problems, unfortunately they are unavoidable. It is only because they fail to understand what it is to have a language that they can think their simulation throws any light on the problem of order. In this respect, the similarity between their "game" and Wittgenstein's builders' "language game" is striking (Wittgenstein 1972). In Wittgenstein's game, the builder points at an object (slab, column) and utters the word "slab", "column" and so forth. As Wittgenstein goes on to point out, whilst we can imagine an individual's vocabulary and indeed a language being enlarged by adding words like this, we cannot imagine a language being initiated in this way. This is because the acquisition of words and their meanings is part and parcel of our social practice of using language itself. Whilst we can have 'private vocabularies', these can only become part of the language and the lexicon by being used as part of the practice language use itself located in an ongoing stream of social life. Thus the builder pointing and naming can only do so because we already have a repertoire of social practices including naming things and other ways of solving problems of reference within a shared scheme of concepts. The notion of a pre-social private language (and hence private set of concepts) which then becomes public and social is simply a conceptual confusion.[30]

Although Marchione et al do not seem to recognise this central problem in their simulation, various definitional decisions taken in their set up allow them to avoid confronting its implications. The game begins with a 'shared' alphabet. What does this mean? The list of symbols to be used is fixed for all participants. Actors can only interpret the words uttered as the 'same' word in their own lexicon (if they have it) and not as a new word spelt differently but pronounced the same, or random noise, ejaculation or whatever. This may seem trivial, but it isn't. The combination of fixing the mapping of word onto object in advance (there is not even the action of pointing as with Wittgenstein's example) and specifying that any interaction can only be about the name of an object defines away the contingency associated with the definition of the situation and the coordination of meaning. Speakers and hearers just do have agreement in meanings.

Whilst the contingency associated with normative compliance has supposedly been set aside, in fact it lurks beneath the surface in the way the form and outcomes of the interactions have been framed. The lexicon is fixed as the combination (4x3!) of a given set of letters. Moreover, hearers have no choice but to add any new word they hear to their own vocabularies. Growth in the lexicon (defined as the set of words in the vocabularies of all actors) is, therefore, enforced. This pair of stipulations imposes normative compliance on the interactions generated by the running of the model.

---

[30] To borrow an example of Quine's. If rabbits are one of the objects in this pre-social, pre-language possible world, and one of these actors utters "gavagai", how does the other know that what is meant is 'rabbit' and not 'undetached rabbit parts', 'member of the genus leporidae' or 'excellent filling for a pie'?

Salgado and Gilbert suggest that the import of the Marchione et al simulation is that the...

> *...results indicate that, when the agents confront an uncertain situation of "double contingency"—that is, they never have direct access to each other's meanings or ontologies—a shared lexicon can emerge, on the condition that a group of agents develops a communicative strategy that favours their mutual understanding and allows them to reach more recipients for their utterances. (Salgado and Gilbert Op. Cit. p 19-20)*

whilst Marchione et al themselves say

> *By using this simulation we have clarified what agent behaviour most effectively spreads their own cultural signs across the population and by a further analysis we have explained why that is the case.*
>
> *The model shows that the group of agents able to reach more hearers and less prone to changing the topic has the highest likelihood of affecting the shared lexicon.(Marchione et al (Op Cit p 13).*

As we have just suggested, it is only because their model assumes mutual understanding and enforces normative compliance that it manages to produce the results it does. In that sense, it is not a strong test of their own success in achieving their ambitions. In as much as these stipulations define away the double contingency at the heart of the classic problem of order as characterised by Parsons and re-interpreted by Luhmann, interesting though it might be technically, in the end the simulation and its 'chatty game' are, unfortunately, of scant sociological interest.

10.4 Analytic Sociology and Agent-Based Models.

In his discussion of mathematical explanation, Alan Baker (2012) makes the point that in science there are mathematically driven mathematical explanations of scientific phenomena and scientifically driven mathematical explanations. The difference between the two is the purpose of the mathematics. In mathematically driven explanations, some physical phenomenon illustrates the mathematical problem being discussed or acts as a hook for deriving and proving a solution to that problem. In the scientifically driven explanations, the mathematics is used in the service of the science. His examples for these two are rather neat. In the former, we ask 'Are hexagons the optimal way of tiling any shape?'. In the latter, we ask 'Why do bees use hexagons?' The answer to the first is a general theorem about hexagons; the answer to the second an account of why hexagons work in Euclidean space.

The same distinction can be used to it to discuss the function of ABM and computational modeling more generally in sociology. There are sociological problems that can be used to explore computational problems, for example 'How can a denumerably infinite language be learned'? And then there are sociological problems that appeal to those who want to use computational methods and solutions on them. Simulation and ABM fall into the latter category. Although there may well be technical interest in the mapping of the computation to the problem (Are the right computations being used and are they being used properly?) the real test comes in the traction that the approach provides on the sociological problem taken up and the insightfulness of its conclusions. From the two examples we have discussed, it would seem that simulation and ABM have still some way to go before we can conclude that it has successfully passed this test. Although it is perfectly acceptable for researchers to want to develop modelling approaches which produce the equivalent of the inclined plane in physics or even the operation of markets in economics, neither of the simulations we have examined has taken us any closer to achieving this goal. Neither Bleda and Shackley nor Marchione et al cast any more light on the sociological problem they focused on nor do they provide any further traction in resolving it. What they have done is translated the problem into a computational formalisation or formulation and then explored the consequences of those representations. Once the model is in place, the sociology gets left behind.

What does all this mean for AS? From the examples we have examined, it seems that although when deployed carefully, simulation and modelling may throw light on the interesting problems of the formalisation of informal and discursive sociological theory, we cannot as yet be confident that they will have much to offer the development of the body of theory itself. Either, as with Bleda and Shackley, the theory has to be re-configured to make modelling tractable or, as with Marchione et al, the key questions at issue have to be assumed away in order to allow the simulation to operate. In addition, the vexed question of empirical validity remains unaddressed. None of the examples we have discussed (neither the two just mentioned nor those in the *Handbook*) can claim any kind of empirical validity for the models they present or the results they generate. The consequence of this must surely be a confirmation of our earlier somewhat negative conclusion. If Analytical Sociology wants to provide an empirically grounded methodology for the development of middle range theories which will demonstrate how macro social structures emerge from micro ones, as they are currently constituted, simulation and agent-based modeling are unlikely to be of much help.