The Methodology of Analytical Sociology

The challenge of creating a robust empirical sociology.

R. J. Anderson & W. W. Sharrock

THE METHODOLOGY OF ANALYTICAL SOCIOLOGY:

THE CHALLENGE OF CREATING A ROBUST EMPIRICAL SOCIAL SCIENCE

R.J. Anderson
University of Nottingham

W.W. Sharrock
University of Manchester

Table of Contents

Preface i

Part A : The programme of Analytical Sociology 1

1.0 The current state of sociology and the objective of analytical sociology 1

2.0 Theoretical Principles 5

3.0 Theories of the Middle Range 12

Part B: Causes and Mechanisms 17

4.0 Explanations and causes 17

5.0 Mechanisms, mappings and metaphors 45

Part C: Agent-Based Models 49

6.0 Agent-Based Modelling as the Method of Analytic Sociology 49

7.0 Cultural Theory and Formalisation 58

8.0 COWCULT and The Formalisation of CT 71

9.0 COWCULT, CT and Formalisation 80

10.0 Agent-based models as sociological analysis 83

Part D: Swans and Ducklings or leopards and Spots? 90

Bibliography 98

PREFACE

As a form of study of human behaviour, sociology has been around for at least 150 years; probably a great deal longer if one is prepared to be relatively open-minded about what counts as sociological reflection on social life. It is, then, a bit long in the tooth for it to be continually excusing or explaining away its lack of empirical rigour on the grounds of its immaturity. That it has not developed a disciplinary structure which replicates or matches those of the natural and mathematical sciences cannot be because it has not had enough time to do so, nor that enough effort has not been expended on the task. Time and time again, figures in the discipline have pushed themselves forward, aspiring to be sociology's Galileo, Newton or Darwin (or, perhaps more likely, its Roger Bacon and Gilbert White) and yet still it remains recalcitrantly unwilling or unable to arrive at any consensus around an empirically based, formalised method. Despite all efforts to the contrary, its theories, methods, data and results remain 'soft' when compared to those of 'proper' science.[1]

In this monograph, we look at the latest attempt to dragoon sociology into a format which can stand comparison with the natural sciences. Analytical Sociology (AS) proposes to do this by re-building the discipline around theories of the middle range designed to offer explanations of the causal mechanisms which generate social structures. These explanations will resolve, once and for all, the two central challenges facing sociology; how to account for the relationship between micro and macro social structures and how to do so by documenting the mechanisms involved in an empirically grounded way. AS is, then, a composite bundle of philosophical principles, a general framework and directives for investigative methods. We will examine all three, and will refer to the bundle as Analytical Sociology's 'methodology'. One of the things which makes AS distinctive[2] is its wish to adopt agent-based modelling as an investigative technique. We will also spend some considerable time examining this proposal. Having reviewed the case which Analytical Sociology makes for its proposed re-direction, our conclusion is likely to be unwelcome to its proponents. As far as we can see, the serious problems involved in grounding an empirically based, causal social science have not been overcome. The central puzzles and dilemmas remain. Moreover, the turn to agent-based models and other forms of simulation is likely to add yet more confusion and unresolved problems to the *potpourri* we already have. In the last section, we reflect on some of the broader cultural and professional factors which seem to be

[1] The contrast between 'hard' and 'soft' disciplines seems much less prevalent in commentaries on the social sciences these days. We would like to believe that this was because of the realisation that the social sciences' problems are just as hard, but differently so, as those of the natural sciences. Alas, we suspect it has more to do with the waning interest in and influence of the social sciences, and sociology in particular, among public intellectuals and policy pundits.

[2] Possibly the only thing, given that it is based on recycling very traditional positions.

preventing this perennial search for new beginnings, new starts and re-organisation from ever being successful. In our view, those who wish to drive sociology closer and closer to the natural sciences might do well to think more about its organisational character than its epistemology before formulating their campaigns.

A number of people have been helpful in the preparation of this monograph. We would particularly like to thank Mercedes Bleda and Simon Shackley for giving us access to the detail of their STELLA model of Cultural Theory which we discuss in Section C. Leonidas Tsilipakos was inadvertently instrumental in prompting us to undertake this study and we thank him for this stimulus. We would also like to thank Stuart Reeves and Murray Goulden for their help and the Horizon Digital Economy Research Institute for its support.

PART A : THE PROGRAMME OF ANALYTICAL SOCIOLOGY

1.0 THE CURRENT STATE OF SOCIOLOGY AND THE OBJECTIVE OF ANALYTICAL SOCIOLOGY

1.1 Same old same old.....

In 1979, Hubert Blalock used his Presidential Address to the American Sociological Association (Blalock 1979) to launch a jeremiad against the sociological profession of his day for its failure to address fundamental questions of conceptualisation and measurement. Rather than take these tricky and complex issues on, the profession seemed to be deliberately sidestepping and ignoring them. This had led to a situation where

> *....in many respects we seem badly divided into a myriad of theoretical and methodological schools that tend to oversimplify each other's positions, that fail to make careful conceptual distinctions, and that encourage partisan attacks. (Blalock op cit p. 881)*

Using Robert Merton as his exemplar, Blalock's prescription was for the discipline to recognise the trouble it was in, to accept the challenge in undertaking the serious work of detailed conceptual and methodological development and, coming together with a sense of common purpose, step by slow step, gradually put sociology on a firm, empirically and conceptual robust footing. If it did not do these things, Blalock warned

> *......I fear that sociology in the year 2000 will be no more advanced than it is today, though perhaps it will contain far more specializations, theoretical schools, methodological cults, and interest groups than, even today, we can readily imagine. (ibid p 894).*

Almost 20 years later, James A. Davis used the same platform to offer his own equally scathing analysis.

> *What is wrong is that Sociology is incoherent. It does not cohere ("to stick together; be united; hold fast, as parts of the same mass"). While each article/book/course may be well crafted, they have little or nothing to do with each other. They may share methods and even data sets (and grammatical voices so passive as to suggest a drug problem), but each is about a unique problem with a unique set of variables.*
>
> *Try this test: list the key concepts/variables in each article of in the last two or three issues of the American Journal of Sociology, American Sociological Review, or Social Forces. I expect the number of different variables will be at least 20 times the*

> *number of articles and few variables (save for a handful of demographics such age, sex and race) will turn up in more than one article.*
>
> *Another indicator: List the major subfields of sociology. Then try to arrange them in some pattern that has more intellectual bite than alphabetization. Hard, isn't it?*
>
> *Yet another: Why are there no conflicts over priority in Sociology? Because sociologists are nice? Nope. Because no two sociologists ever study the same thing, so such conflicts are impossible (Davis, 1994, p 180).*

The year 2000 has come and long gone and no doubt, were he still alive, Blalock would feel his warnings have gone unheeded and his gloomy prognostications realised. Certainly there is no doubt that, to-day, Davis could repeat his charge word for word since so little has changed in the meanwhile. The distinguishing feature of sociology remains a deep lack of consensus on how to make progress in the discipline. No theoretical constructions have universal or even near universal support. There is no consensus on the methodological programmes definitive of the research practice of the discipline. No broad based cadre of studies generating a cumulative body of knowledge is to be found. All around is difference and dissent.

By and large, responses to this condition take the same forms they always have. Either there is a Tybalt-like cursing of all dominant modes or, as with Blalock and Davis, there are strident exhortations to adopt the *modus operandi* of synthesis and integration and rally around some favoured version of one of the current contending schools.

1.2 Analytical Sociology to the rescue?

AS shares the views of Blalock and Davis, though it tends to be somewhat less condemnatory in its phrasing. It too sees sociology exhibiting disarray. Here are the opening words of *Dissecting the Social,* Peter Hedstrom's (2005) initial survey of the possibility which AS offers.

> *Over the past several decades leading sociologists in Europe and in the United States have expressed strong reservations about the explanatory power of sociological theory and research ....They are concerned that much sociological theory has evolved into a form of metatheorizing without any specific empirical referents, and that much empirical sociological research has developed into a rather shallow form of variable analysis with only limited explanatory power. The main message of this book is that a path must be hewn between the eclectic empiricism of variable based sociology and the often vacuous writings of the 'grand' social theorists. (Op. Cit. p 1.)*

Hedstrom is very clear how this path is to be hewn out.

> *(T)he advancement of social theory calls for an analytical approach that systematically seeks to explicate the social mechanisms that generate and explain observed associations between events......In the case of sociology.....a sustained focus on explanatory social mechanisms would allow sociological theory to reconnect with what we consider to be its most promising and productive era —*

> *namely, middle range sociology of the kind that Robert Merton and Paul Lazarsfeldt tried to develop....(Hedstrom & Swedberg 1998 p. 1)*

Middle range theorising is to be the salvation of sociology. Yet this is no new and different modality but a synthesis of the best of what we already have (which to judge from the quotation above, is not much). The aim is to consolidate around tested principles and facilitate piecemeal development in theory and explanation. This unification will take place in two ways. There will be a unification of theory and investigation. Theory will become empirically grounded. There will also be a synthesis within theory, with the rival theoretical frameworks being integrated into a unified field. Pierre Demeulenaere puts it like this.

> *Analytical sociology should not therefore be seen as a manifesto for one particular way of doing sociology as compared with others, but as an effort to clarify ("analytically") theoretical and epistemological principles which underlie any satisfactory way of doing sociology (and, in fact, any sociology)....... The aim of analytical sociology is to clarify the basic epistemological, theoretical and methodological principles fundamental to the development of sound description and explanation. (Demeulenaere 2011 p. 1)*

These are challenging objectives, albeit expressed somewhat *sotto voce*. Achieving the clarity of principles Demeulenaere is asking for and then implementing them as an operationalisable methodology for sociology will be neither easy nor straightforward. But only if AS can achieve the latter, will the value of the former be realised.

In this monograph, we will look at the challenges AS has set itself and the way that it has set about overcoming them. Our focus will be on the robustness of the principles which have been and are being adopted and the feasibility of basing an empirically grounded investigative programme on them. We will also look at some of the studies claiming to demonstrate this possibility. Our purpose is to examine the *methodology* of AS. Because this term has come to have a somewhat different usage in sociology to that which we intend, it is worth spending a few moments at the start explaining what we mean by it.

1.3 Methodology and the logic of sociology

We want to distinguish between methodology and method; or between methodology and methodologies. Following Felix Kaufmann (1958), we take methodology to be the study of the logic of an investigative discipline. Methods or methodologies are the techniques which might be employed to undertake such investigations. Investigation might, as in logic or mathematics, take the form of testing of the validity of a set of inferences derived from a set of premises and axioms. Or it might, as with empirically driven disciplines, be the testing of theoretical propositions against evidence. We are not here concerned with philosophical questions concerning the security of deductive or inductive inference. Nor are we concerned to legislate that sociology should only conform to one or other. Philosophers of science and sociology have long struggled with these issues. For the moment, we simply acknowledge the (seemingly unending) debates and take it as a matter of fact[3] that sociology aspires to be a body of theory and empirical investigations. The methodology of sociology is simply how that theory and those investigations are aligned so that the theory actually motivates the investigations and the investigations actually provide evidence (positive or negative) for the theory. Of course, 'alignment' is a weasel-word. We sociologists are adept at finding some connection, some alignment between a theory and an investigation whose results we want to argue are relevant to the theory. So specifying *some* connection is not enough. The diatribes of Blalock, Davis are really about just how easy it is in sociology to make some connection between a set of findings and whatever theory we like. For sociology to be a robust discipline, its methodology must contain requirements that rule out loose connectivity. There must be clear, defensible and strict rules by which we can step from theory to investigation to findings and back again. Those rules will specify the logic being followed. If AS wishes to be successful in, first, achieving the integrations that it has set its sights on and, second, re-starting progress towards constructing a discipline that can stand comparison to the physical and natural sciences, then it will have to forge its own methodology.

Adapting some of James Woodward's (2003) recommendations, we suggest that a robust methodology for an investigative sociology should demonstrate the following virtues:

1. The investigations carried out under the methodology should provide adequate descriptions; that is, capture the paradigmatic features of social life which members of the society experience.

2. If the sociology proposes to integrate a variety of alternative theories, it should show how that integration is to be achieved whilst preserving the integrity of each theory and the coherence of the whole.

3. The methodology adopted should allow us to evaluate which investigations and explanations are effective and which ineffective, which good and which poor. One of the anxieties noted by Schweder and Fiske (1983) over a quarter of a century ago was the discipline's inability to enforce standards, not just over quality but of how to distinguish non-science and nonsense from genuine contribution.

4. The methodology must have secure enough epistemological underpinnings to license the investigative approaches adopted. That is, the methodology must provide for robust and effective investigative methods.

Our question in this monograph is simple: 'How far does AS' methodology satisfy these criteria?' Although this is a question of prime importance, a positive answer may only be available over the long term. This is because, as well as pulling theorising together into a common framework and marshalling investigative methods so that the findings of studies will cumulate, AS has to find a way of preventing itself being paralysed by having to deal with arguments in the philosophy of the social sciences. To do that, it will have to find a way of either

[3] Whatever *that* means!

answering those arguments or of generating the confidence to ignore them.[4] A negative answer, however, may be arrived at in shorter order and, if it is, might well encourage AS to reflect on what else it needs to do to realise its goals.

2.0 Theoretical Principles

2.1 The metaphysics of agency and structure

The objects of sociological enquiry, its phenomena, are the structures and institutions produced by patterns of social action. These structures and institutions are the intended and unintended consequences of such patterns. In so far as the behaviour of individuals is a topic for sociology, it is always set within some micro or macro patterning. Typical examples of macro structures as such things as the economy or the state. For micro structures, they are face to face groups, individual families, work teams and dyads. It is important to understand that the macro-micro distinction is a definitional binary not the end states of a spectrum or continuum of structures and organisations where some might be 'macro cum micro' or 'micro cum macro'. The ontology of sociology is, then, one of individuals and structures where the latter are either macro or micro. This ontology generally appears as the explanatory pairing of structure and agency.

The agency/structure pairing is the conventional metaphysics for most sociological theory. Modes of theorising are defined in terms of where they place explanatory weight. True, some, do seek to reconcile individuals and structures (as with Giddens' (1984) notorious theory of "structuration" and, as we will see in a moment, with AS' middle range theorising). But by and large, the texts and the summaries of theory present sociology as divided. Whilst AS (and Giddens) want to synthesise bi-polar explanations, they do not want to reject the ontology they rely on. In the sociological world there are only macro and the micro structures and individuals. When individuals do feature in the sociological world, they do so as members of micro structures and/or macro structures. They are individuals-in-a-structure not individuals *qua* individuals. Thus explanations which invoke individuals are explanations in terms of social-individuals-in-a- social-structure. This corollary of the macro-micro stipulation, that individuals are individuals-in-a-social-structure is what, for most sociology at least, prevents explanations in terms of individual agency collapsing into psychological explanations in terms of the desires, wants and needs of individuals. This axiom of sociality will become important later in our discussion.

AS does not question the ontology of the individual-macro/micro pair. But it does see that accepting it could force an explanatory choice. AS tries to avoid making that choice by adopting the twin principles of 'structural individualism' and 'supervenience'. Structural individualism is the *methodological principle* that all explanations in sociology are to be couched solely in terms of the agency of individual social actors. There is only one source of causal efficacy for sociological phenomena and that is the action of individuals. It is a version of the broader principle of methodological individualism. Supervenience is an attempt to avoid the

[4] It was, we think, Murray Gell-Man who observed that he could think of no advance in the physical sciences that had been helped (let alone stimulated) by debates in the philosophy of science.

explanatory difficulties faced by theories that want to invoke linked conceptual bundles such as 'individual self interest' and the institution of a 'social contract' or 'rational choice' and the operation of 'the hidden hand of markets' in order to allow structural patterns to emerge from individual action but not be caused by them. The relationship between micro-structures and macro-structures is not causal but constitutive or emergent. The combination of the two principles is designed to place joint explanatory weight across both macro and micro without being reductionist or causalist.

It is crucial to be clear what AS is saying here. The only causal agents in social life are individuals.[5] Individuals act together in micro structures. These structures can bring about (causal) changes in other micro structures. Such micro-foundations constitute macro-structures as parts to a whole. As a consequence, the chain of causal effects at the micro level produces supervenient changes in the macro order. This is what is called the "Coleman boat" in the AS and related literature.[6] The causal path of change in the macro order is through the micro order which is then articulated at the macro level. This is captured by the following diagram taken from Hedstrom & Bearman's (2009) introduction to the *Handbook of Analytical Sociology.*

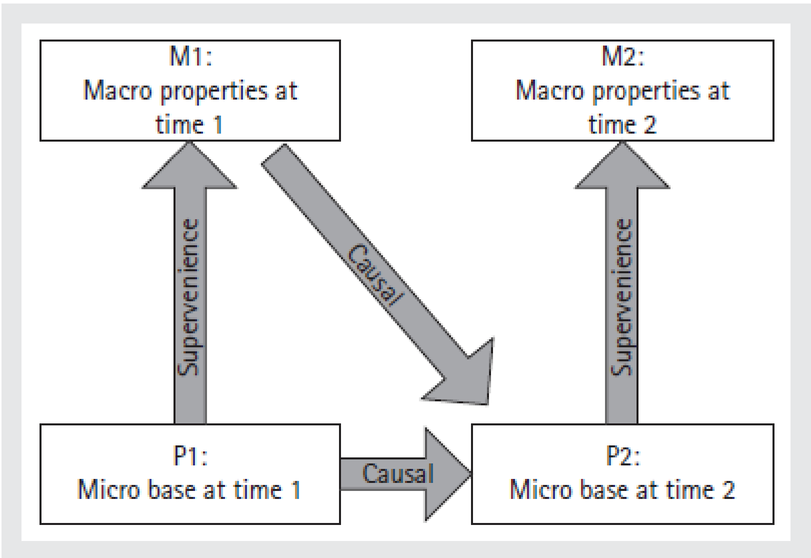


Fig. 1.2 Macro dynamics from a supervenience perspective

2.2 Epistemology and social reality

When we look around the social world, we don't 'see' structures and organisations. We see people doing things, either on their own or with others — driving cars, buying jewelry, going to school, talking to their children. For AS, any causal or other sociological explanation of these doings cannot simply be a listing of who did what, when and where. To cast explanations at this level would be to embrace crude empiricism. AS insists

[5] It will be important to understand what AS means by 'individuals' and what is entailed in tracing causes at an individual level. Given the complexity of social life in even the simplest societies, tracing the causal paths at the level of particular individuals will be quite a challenge. If AS means 'individual-as-a-social-type' then we have to confront the methodological question of how we go from actual individual actions to the actions of (ideal?) types.

[6] This wasn't actually invented by Coleman. See Barbera (2006) fn15

that through a process of abstraction we must dissect the complex totality of social experience by abstracting out immaterial or irrelevant features. This is what theories do. They reduce and abstract.

> *Developing explanatory theory involves a delicate balance between realism and abstraction. Although it is difficult to specify a priori what should be considered a sufficiently faithful representation of a social process, the question is of fundamental importance. Explanatory theories can never be based on fictitious accounts, because such accounts cannot provide convincing answers to the question of why we observe what we observe. What must be aimed for is 'analytical realism'...... (Hedstrom 2005 p 3)*

The term 'analytic realism' was coined by Talcott Parsons. Here is how Parsons defines this position.

> *....it is maintained that at least some of the general concepts of science are not fictional but adequately "grasp" aspects of the objective external world.....hence the position here taken is, in an epistemological sense, realistic......These concepts correspond not to concrete phenomena, but to elements in them which are analytically separable from other elements. There is no implication that the value of any one element, or even of all of those included in any one logically coherent system is completely descriptive of any concrete thing or event. (Parsons. 1949 p 730)*

Theoretical terms are organising categories which enable sociology to analyse social life through a process of abstraction. Since it is not possible to carry out analysis without such categories, they are apodictic. The mapping between the concept and the social world is taken as given. We cannot interrogate the social world to see if there are structures and organisations. Rather, the presumption of structures and organisations is what enables us to provide sociological explanations.

This is not the place to tease apart Parsons' notion of analytic realism and the interpretation of Kant on which it is based. Suffice it to say, first, that it is far from accepted that, despite his heroic efforts, Kant actually did solve the problems of empiricism; and, second, that the version of Kant which Parsons calls upon will actually stand the strain he puts it under. However, having adopted analytic realism, AS has quite clearly set itself a challenge. The process of dissection and abstraction runs two risks: over-abstraction and a loss of groundedness in the detail of social reality on the one hand, and on the other a lack of generalisability because of the account is overly grounded in empirical detail. How much abstraction should we aim for? How much detail do we need? The challenge to AS is how to determine how to answer these questions.

2.3 Structural Individualism and semantic symmetry

We can readily enough grant AS the right to choose its own premises. We can happily allow it to proceed *on the basis that* theories must aspire to be "analytically real" and all explanations of social facts must be couched in terms of the actions of individuals, and then see what that standpoint delivers. However, if the principles are to be *principles*, they must be applied in a coherent and consistent manner, and the difficulties which they might generate addressed. For structural individualism, the problems arise in the rules for re-writing descriptions of macro phenomena (societies, groups, markets etc) first into micro phenomena (families, teams, groups) and then into descriptions of individual action which can be re-transformed back into explanations of aggregate phenomena. This two-way path is needed because AS wants to aggregate individual actions into social structures and thereby gain explanatory access to effects which are invisible when viewed from the individual point of view. To use one of Thomas Schelling's examples (see below), summing over the preferences of all individuals for housing choice, you do not find racial discrimination. However, over time, extreme racial segregation emerges from the actions of all the individuals making decisions based upon their own preferences. The question is not whether the processes in any particular housing area could have been different. There is no way to run the time line backwards to get to the original state because of the *other* consequences which follow from the emerging pattern of behaviour. These are such consequences as changes in employment patterns, education, value of housing stock and so on. This, then, raises the question whether you can re-write descriptions of the phenomena observed at the aggregate level into the disaggregated level *and retain the properties/characteristics* which you are trying to explain, or at least whether you can do this in a way which is reasonably plausible for an actual case. To take another example we will look at in detail later, that of romantic attachment among adolescents (Bearman et al 2004). The spanning tree network describing these relationships is (only??) rationalisable in terms a social norm which says 'Don't go out with the prior girlfriend of your prior girlfriend's current boyfriend'. However, summing over all expressed preferences for their romantic partners of the young people, this formulation is invisible. Moreover, it doesn't decompose into any of the norms which they were actually orienting to. How then is its descriptive, let alone explanatory, basis to be grounded? At what level of meaning are we going to locate its *explanatory force*?

2.4 Emergence, the synechdoche problem and the mystery of supervenience

Emergence is an ontological category not an explanatory one. Properties are observable (or emerge) at one analytic level but are not observable at or reducible to another, usually lower, one. The familiar examples are the translucence and liquidity of water between 32° C and 100° C. The molecules of hydrogen and oxygen which make up water are not translucent or liquid within this temperature range. A description of the properties of the separate gas molecules will not be a description of the properties of water as a combination of the molecules. In sociology, it is conventional to define social institutions, structures and organisations as emergent from the actions of individuals and groups.

Philosophers use 'supervenience' to explain the relationships between orders of emergent properties and their constituent elements without needing to invoke causation. For example, discussions of the relationship between mental properties or states and physical properties and states, suggest that mental properties supervene on physical ones but are not caused by them. Or again in regard to moral discussions, moral properties (being right, being better) are held to supervene on natural ones (being human, or dense or made of green cheese). In both cases, supervenience only points to necessary correlation: "No A properties without B properties"; no mental properties without physical ones; no moral properties without natural ones. If we are not to take supervenience to be a causal relationship of some kind (that is, if 'to constitute' is not to be taken as a causal verb), just how do causally deep explanations couched in terms of structural individualism account for supervenient relationships?

In the case of the mental and the physical, the proponents of supervenience argue that while specific states and processes in the brain are necessary for us to have memories, thoughts, weigh options and so on, those states and processes do not determine our memories, thoughts and choices. This disjunction is required for there to be philosophically defensible concept of personal agency. However, whilst it offers a relatively neat solution to the problem of how to relate the concepts of the mental and the physical, it does so at the price of leaving the actual, empirical connection between them a mystery.

AS wants to say that macrostructures supervene on micro ones; microstructures are necessary for macrostructures. This is an assertion about the correlation of two (analytically real) categories. Here is how Hedstrom and Bearman describe supervenience between micro and macro structures.

> *.....a macro property, $\boldsymbol{M}$, supervenes on a set of micro-level properties, $\boldsymbol{P}$, if identity in $\boldsymbol{P}$ necessarily implies identity in $\boldsymbol{M}$. If the macro property is supervenient upon the micro it means that, if two collectivities or societies are identical to one another in terms of their micro-level properties, then their macro-level properties also will be identical. It also implies that two collectivities that differ in their macro-level properties will necessarily differ in their micro-level properties as well. But it does* not *imply that two collectivities with an identical macro-level property will necessarily have identical microlevel properties, because identical macro-level properties can be brought about in different ways.*
>
> *Although macro is dependent upon micro, micro-to-macro or $\boldsymbol{P}$-to-$\boldsymbol{M}$ relations should not be viewed as causal relations. Macro properties are always instantiated at the same time as the micro properties upon which they supervene, and a group or a society has the macro properties it has in virtue of the properties and relations of its micro-level entities. The micro-to-macro relationship is a parts-to-a-whole relationship rather than cause-to-an-effect relationship. For example, if a set of dyadic relations exists between the members of a group, these dyadic relations do not cause the network structure linking the individuals to one another; they constitute it. Similarly, the properties of the individuals residing in different spatial locations do not cause the extent of residential segregation; they constitute it. (Hedstrom and Bearman 2009 pp10-11)*

We are now faced with two problems. First, given explanatory supervenience, what does the objective of offering explanations with "causal depth" mean? Hedstrom and Bearman define causal depth as:

> *By causal depth we mean the explicit identification of the microfoundations, or the social cogs and wheels, through which the social facts to be explained are brought about. The central cogs and wheels of social life are actions and relations. Actions are important because all the things that interest us as sociologists are the intended or unintended outcomes of individuals' actions. Individuals' actions typically are oriented towards others, and therefore relations to others are central when it comes to explaining why individuals do what they do. In addition.... social*

> *relations are central for explaining why, acting as they do, individuals bring about the social outcomes they do. That relations are important for explaining outcomes does not mean that they are independent of individuals and their actions, however. As emphasized above, in principle all relational structures are explainable as intended or unintended outcomes of individuals' action.......*
>
> *Causal depth is achieved by recognizing that action takes place in social structures that in this case channel mobility opportunities and thereby explain why we observe what we observe. (ibid p 9)*

If we cannot say that the properties of macro structures are caused by the properties of micro structures, what does it mean to say that we want to identify the microfoundations through which social facts observable at the macro level "are brought about"? Moreover, in that they talk about these explanations as causally deep and not 'superveniently deep' ones, presumably we are to assume that Hedstrom and Bearman want us to think of this "bringing about" as causal.

The second issue has to do with what is called the synechdoche problem; how to separate the meaning of constitutive parts and wholes. If we are to divide a pie among four children, the four quarters constitute the pie. Ontologically, do we have four pieces of pie *and* a pie in four pieces? While the pieces do constitute the pie, each piece only gets its sense and identity from being part of the pie. How do we reason about the parts without reasoning about the whole? And how do we reason about the whole without assuming the parts? We know which AS says has explanatory priority but which is to have *ontological* priority? If AS wants to explain the deliberations of a political structure in terms of the actions beliefs and desires of a set of political actors, how do we make sense of (and hence investigate) political actors without seeing them as part of the relevant political structure? But, for AS, that structure is constituted by the actions of those actors. Of course, we could propose, as Hedstrom and Bearman do, that the relationships are temporal and iterative. One set of actions, or the structure they generate, causes another set of actions which then constitute a changed political structure which then has its own causal consequences, and so on. . But that does not solve the problem of explaining synchronous micro and macro structures. For these, the synedoche problem remains and all that supervenience gives us are explanatory race conditions.[7]

Of course, we might solve both problems by adopting sociology's usual trick for getting out of tight analytical corners, namely the 'point of view' point of view. Looking from the point of view of macrostructures, institutions, organisations and structures supervene on microstructures. Looking from the point of view of microstructures, such institutions etc. are caused by the actions of individuals through microstructures. Point of view hopping is an attractive and widely used tactic. Unfortunately, it does not make for theoretical integration, coherence and consistency.

[7] What supervernience does solve (or rather sidestep) is the *ontogenesis* problem. We do not have to work out how we can create aggregate social life out of individual asocial actions. That is, how we go from 'no social life' to 'social life' through the actions of individuals.

Given these difficulties, perhaps some light might be shed if we look at the modality of supervenience relationships. What does "necessarily" mean in the assertion "A properties necessarily co-vary with B properties"? The property bundle A necessarily co-varies with the property bundle B in as much that there can be no change in A without change in B. How are we supposed to take this? There appear to be three alternatives:

- Perhaps A and B are identical - that is, A is B? For instance, H_2O is water (at least in our possible world) and water is necessarily H_2O. The properties of water (boils at 100c, freezes at 0c, is translucent, etc etc) are identical with the properties of H_2O. Is this what AS is claiming for macro and microstructures? It would seem not, for if they were identical there would be no constitutive or causal relationship to explain.
- Perhaps A is logically entailed by B? Being A is part of the meaning of B. Being a bishop supervenes on the attribute of being a cleric. But if the characteristics of some phenomena, say a run on the stock market, are logically entailed by the properties of the actions of individuals, then isn't it entailed by the meaning of the term 'run on the stock market' that it means the specific actions of individuals? Is this what AS is claiming; that is that the meaning of any particular macro-structural term is that it is or is not part of some specific micro-structural term? If this is so, then the relationships are conceptual. The trouble with this solution is that conceptual relationships are not amenable to causal explanation.
- Perhaps A is metaphysically contingent on B? This does open up the space for causality. Take Boyle's Law: $\frac{PV}{T} = k$. Temperature, pressure and volume supervene on each other. No change in one without change in the others. This relationship is nomological in our world but, of course, there are possible worlds (and, perhaps, possible set up conditions in our world) in which Boyle's Law might not apply. In those worlds, Boyle' Law, where it applied, would be an inductive generalisation. In that Hedstrom and Bearman argue that the same macrostructure might be constituted by two different microstructures, it looks as if supervenience is being conceived in terms of metaphysical contingency. In the social structures we know about and can imagine, macrostructure A can supervene upon microstructure B or upon microstructure C. This co-variation is not nomological (the two do not co-vary like mass and gravity), but is an empirical generalisation. Even if all the macrostructures of type A we have studied supervene on B, we can envisage A without B but with C, just as we can envisage mammals without backbones though we have never found one. However, because the generalisations sociology can offer (so far at least) are so weak, all we can say is that if A supervenes on B, it does so to some value of $p \leq 1$, and usually much less than 1. This would hardly be the kind of strong generalisation AS wants middle range theory to facilitate.

The key question, however, is not whether we can distinguish and then relate general categories or types but whether we can relate particular instances of them; *which* macrostructures supervene upon *which* microstructures? In biology, it seems we have a reasonable idea about which brain states and process are involved with which mental function. In sociology , do we have any idea how to fix the corresponding set of microstructures B for a particular macrostructure properties A?[8] If we did know how to specify the co-variations (either deterministically, A always occurs with B, or probabilistically, A usually/sometimes occurs with B) then the way would be clear to study the (supervenient) relationships between A and B. At the moment we don't. To bridge this investigative lacuna, AS has recently turned to agent-based modeling (ABM). AS expects ABM to specify the relationships between particular microstructures and particular macrostructures and so dispel the mystery. As we will see, this might well be a vain hope.

3.0 Theories of the Middle Range

3.1 The Problem

The analytic pull of AS is ecumenical and centroid seeking. It values incorporation and convergence. We have already seen this tendency in the way that causation and supervenience are brought together. It is also very clearly visible in the scope of sociological theorising which AS prefers— namely what Robert Merton called theories of the middle range. For Hedstrom and Udehn, middle range theory is

> *.....a clear, precise and simple type of theory which can be used for partially explaining a range of different phenomena, but which makes no pretense at being able to explain all social phenomena, and which is not founded upon any extreme reductionism in terms of its explanans. It is a vision of sociological theory as a toolbox of semigeneral theories. In this sense (the) vision has more in common with the type of theories found in the life sciences than those found in the physical sciences. (Hedstrom and Udehn 2009 p 31)*

The notion of middle range theory is explicated by mapping sociological theory along two dimensions; the generality or particularity of the *explanadum* and the inclusivity or exclusivity of the *explanans*. This allows Hedstrom and Udehn to produce a conventional 2x2 matrix with the examples given below occupying the definitional polar cells.

[8] As we have just seen, AS argues that different microstructures can produce the same macrostructure. Microstructure A is sufficient for macrostructure B but so is microstructure C. Neither A nor C is necessary but *some* microstructure is. Quite how this fits into the pattern of causal explanations in terms of sufficient and necessary conditions is a question we will have to return to.

	Particular	**General**
Inclusive	Geertz	Parsons & Luhmann
Exclusive		Becker and Homans

Talcott Parsons and Niklas Luhmann are said to offer theories which utilise a wide range of explanatory devices and explain the operations of society as a whole. Gary Becker and George Homans, though they are equally general in scope, concentrate on a limited range of explanatory factors. So much, so straightforward (if not uncontentious). The problems come with the *explanadum* dimension. Hedstrom and Udehn designate Clifford Geetz' notion of "thick description" as an example of inclusive/particular theories. This must mean something like the description of a particular pattern of action by enumerating the contextual detail in which that pattern is to be found. Certainly Geertz is concerned with rendering contextual detail, but this is to underpin ethnography as an interpretive method. His interpretive schema are as general as those of Parsons and Luhmann, Becker and Homans. It is just that they are tied to the detail of cases. At the other end of this dimension we have descriptions which are highly focused and low in scope, i.e. idiosyncratic accounts of singular phenomena. Hedstrom and Udehn call them "thin descriptions" but that is hardly illuminating. We can only think that such descriptions would be decontextualised summaries. It is hard to imagine anyone in sociology with that narrow and abstract a descriptive horizon. Indeed, it may be impossible to do sociology with that kind of horizon, which may explain why they offer no examples to illustrate what they mean.

Since it is so loose, the framework Hedstrom and Udehn use is not really that helpful as a mapping of sociological theory and therefore for defining the problem for which middle range theory is being offered as the solution. Although one of its dimensions (Particular/General) does looks robust, actually it is univalent. Sociology does not trade in theories which operate at the particular level. Of course, the purpose of the mapping is not really to provide an viable summary of extant sociological practice. Rather, it is to provide a device whereby middle range theory can occupy a unique central place and from which it can offer a unifying strategy. It can be presented as the "Just Balance" of inclusive/exclusive - particular/general theories. The weakness of the initial organising framework is only important analytically because once the it is in place, AS goes on to reduce the two dimensions to the classic macro-micro continuum discussed earlier and to use this as its own theoretical rationale. Integrating explanations of macro and micro (rather than, say, dispensing with the contrast altogether because the framework which licenses it is so weak) is what AS is about.

The positioning given by Hedstrom and Udehn isn't quite how Robert Merton presented middle range theories. For Merton, the contrast between general theory on the one hand and empirically oriented working hypotheses on the other was rhetorical rather than programmatic. What Merton was arguing for was what he called (quoting T. H. Marshall) "stepping stones into the middle distance" to replace the prevalence of theoretical leaping from guesses, surmises and findings about individual cases to grand theoretical systems.

> *Middle-range theory is principally used in sociology to guide empirical inquiry. It is intermediate to general theories of social systems which are too remote from particular classes of social behavior, organization and change to account for what is observed and to those detailed orderly descriptions of particulars that are not generalized at all. (Merton 1968 p 39)*

Merton felt the need for such theory was urgent because of sociology's immaturity. Trying to develop general theory was a wildly optimistic ambition when the required ground work had not yet been done. In an enlightening comparison, he drew a parallel between sociology as he found it and the state of medicine in the 17th century. Both are simply incapable of forming the generalisations required to provide a fully articulated robust general theory.[9] Rather the strategy must

> *......proceed on these interconnected planes: (1) by developing special theories from which to derive hypotheses that can be empirically investigated and (2) by evolving, not suddenly revealing, a progressively more general conceptual scheme that is adequate to consolidate groups of special theories. (ibid p 53)*

3.2 The Solution

For Merton, then, middle range theorising depended on and encouraged the accumulation of tried and tested findings. The required theory was to provide

> *......logically interconnected sets of propositions from which empirical uniformities can be derived. (P 39)*

Merton summarises his proposal for piecemeal theorising in the following way.

> *1. Middle-range theories consist of limited sets of assumptions from which specific hypotheses are logically derived and confirmed by empirical investigation.*
>
> *2. These theories do not remain separate but are consolidated into wider networks of theory......*
>
> *3. These theories are sufficiently abstract to deal with differing spheres of social behavior and social structure, so that they transcend sheer description or empirical generalization.....*
>
> *4. This type of theory cuts across the distinction between micro-sociological problems, ... and macro-sociological problems......*
>
> *5. Total sociological systems of theory......represent general theoretical orientations rather than the rigorous and tight knit systems envisaged in the search for a "unified theory" in physics.*

[9] Interestingly, Merton was a man of his times. His view of the form generalisations should take was that they should be law-like, a position AS now rejects.

> *6. As a result, many theories of the middle range are consonant with a variety of systems of sociological thought.*
>
> *7. Theories of the middle range are typically in direct line of continuity with the work of classical theoretical formulations.....*
>
> *8. The middle-range orientation involves the specification of ignorance. Rather than pretend to knowledge where it is in fact absent, it expressly recognizes what must still be learned in order to lay the foundation for still more knowledge........(ibid. pp 68-9)*

This positions theories of the middle range in a contrast between two ways of demonstrating the utility or validity of general theory. One way (and this is the route chosen by Parsons, Luhmann etc) is to use particular cases to illustrate the applicability of general theories. That is, the particular case is construed in terms of the favoured theoretical terms and so its intelligibility secured. The second way (and this is Merton's route) is to demonstrate the generality of findings couched as theoretical specifications of particular phenomena (for example, relative deprivation or self fulfilling prophecies) by gathering broader and broader collections of cases of them. This is a kind of stepwise induction. Theories of the middle range are such stepwise inductive generalisations. AS adds a further requirement to this strategy. They should describe robust 'social mechanisms'.

Merton didn't think he had invented a new style of sociology. Indeed he is at pains to point to an array of prior and contemporary work which he feels fits the style of work he is pushing for. All he is doing is coining a name for what they produce. What is interesting is that AS does little more than gesture at such precursors.[10] When it sets out what middle range theory and related social mechanisms are supposed to be, it cites just three (or four) of Merton's own worked out examples; structurally constrained opportunities; unanticipated consequences of action; self fulfilling prophecy (and its correlate the Matthew effect). This is surely a somewhat narrow base on which to propose a whole re-direction! There may be an obvious reason for this, of course. Very few of the people Merton cites saw themselves as contributing to the style of theorising he was promoting. He saw them as co-members of his campaign. They did not. Some such as Gouldner were vehemently opposed to it. So drafting them in as contributors to nascent AS is likely to start more hares and generate more problems than it will solve. However, this absence of reference to exemplars does raise the question which sociological studies being carried out today Merton would be inclined to claim for his case and why AS has not tried to capture them for AS too. It is equally interesting to note how few (if any) of the contributions to AS are actually positioned as instances of Merton-type theory. Is that too because they do not see themselves working in this vein?

The above considerations force two questions to the fore:

[10] We are not alone in noticing this absence. Crowthers (2013) makes a similar point.

- To what extent are the mechanisms which are the core of AS' methodology actually inductive generalisations?
- How many of the mechanisms identified in studies actually have any sort of generalisability?

Without strong positive answers to both these questions, it is hard to see how AS can claim Merton's *imprimatur*.

3.3 AS and Middle Range Theory

As we have just seen, Merton was explicitly concerned with development of standard forms of sociological theory. The focus which AS places on mechanisms sits oddly with this. To be sure, in passing Merton does say that such theories will contain explanations of the mechanisms by which social effects are created but his is a casual, vernacular use. Certainly he does not identify mechanisms with middle range theory in the way that AS does. When we turn to examples of AS for insight, we find they have neither the analytic detail of thick descriptions nor the explanatory power of general theory. For the most part, they are descriptions of individual cases or groups of cases and do not appear to be designed to be aggregated. They have a bespoke character rather than general purpose ones that could act as modules in theory construction. There is no inductivity (to invent a terrible phrase) to them. Instead of trying to convince us we should see mechanisms as what Merton had in mind when he talked of theories of the middle range and what, therefore, we in sociology should be investigating, perhaps those who are promoting AS should be trying to convince their colleagues to build their mechanisms to a (small) set of templates which are designed to provide the cumulative findings which can be integrated to produce theories of the middle range.

Merton introduced theories of the middle range as part of an argument he was making about the strategic direction of theorising in sociology. He felt it was likely to lead to disappointment and disaffection. It wasn't that he was against general theory but that he thought it was too soon in the discipline's history for us to expect to be able to do it successfully. His argument did not rest on a binary division of the discipline (although he did recognise there were plenty of oppositions to be found). Rather, it rested on a contrast of theoretical types. This contrast provided the rhetorical space for him to introduce his proposed re-orientation of the discipline's strategy; one which involved significant but not total theoretical deflation. Although he is fairly clear what the general properties of theories of the middle range might be and lists lots of investigations which have made contributions that fit his bill, nowhere does he specify that such theories should be constructed around a single format, and certainly not that of 'explanatory social mechanisms' as AS conceives them. Seeing the introduction of middle range theories as a rhetorical move rather than a programmatic one obviates the need to give the idea more substance than it actually has. It is only if you want to re-direct sociology entirely that you have to construct middle range theory in terms of mechanisms (or something). Merton wasn't trying to do this but rather give some shape to what was already going on.

PART B: CAUSES AND MECHANISMS

4.0 EXPLANATIONS AND CAUSES

4.1 Causal explanations

As far as AS is concerned, an explanation in sociology is the description of the causal mechanism generating a social phenomenon such as a segregated neighbourhood, a run on the bank, a pattern of romantic relationships. We devote the this part of our discussion to this suggestion. In doing so, we will look at what AS takes 'cause' to mean and how far it is possible for sociology to satisfy the requirements for specifying causes in the way intended. Only when we have this clear, will it make sense to look at the notion of 'mechanism' as a definition of what a cause is and whether mechanisms provide good causal explanations in sociology. Whilst we will separate these ideas out in our discussion, AS routinely runs the two together. As a consequence, the notion of cause appears to be shaped by the commitment to mechanism-based explanations and the notion of mechanism appears, equally, to be shaped by the need to provide causal accounts.

To begin with, how does AS conceive 'cause'? Hedstrom and Ylikowski (2010) approvingly cite James Woodward's (2003) analysis even though in the final chapter of that work Woodward expresses severe reservations about 'mechanism' accounts of causation.

> *In Woodward's account, causal claims track relations of counterfactual dependency. They tell us what would have happened to the effect if the cause had been subject to a surgical intervention that would not have affected any other part of the causal structure. One of the novelties of Woodward's theory is its account of causal generalizations in terms of invariances. According to Woodward, the explanatory qualities of a generalization are determined by its ability to tell us about the counterfactual consequences of possible interventions....(Hedstrom & Ylikowski op. cit. p 54)*

By way of comparison, here is how Woodward himself summarises his position.

> *I favour a broad notion of causation according to which, roughly, any explanation which proceeds by showing how an outcome depends (where the dependence in question is not logical or conceptual) on other variables or factors counts as causal. I suggest that the distinguishing feature of causal explanations, so conceived, is that they are explanations that furnish information that is potentially relevant to manipulation and control; they tell us how, if we were able to change the value of one or more variables, we could change the value of other variables. (Woodward 2003, p6)*

A little later on, he puts it this way.

> *...(M)y idea is that one ought to be able to associate with any successful explanation a hypothetical or counterfactual experiment that shows us that and how manipulation of the factors mentioned in the explanation (the* explanans*, as philosophers would call it) would be a way of manipulating or altering the phenomenon explained (the* explanadum*).....(A)n explanation ought to be such that it can be used to answer what I call a* what-if-things-had-been-different question*.....(ibid. 2003 p 11emphasis in the original).*

This interventionist and counterfactual approach can be applied to both token (or single case) and type (or general example) causal explanations (these are Woodward's own terms). The example he gives of the former is the explanation of the mass extinction at the end of the Cretaceous era. According to current theory, this was caused by a massive asteroid impact. The example of the latter is the explanation of the acceleration of a block on an inclined plane as a function of the angle of the slope, the mass of the block and gravity.

We will not examine Woodward's theory of causal explanation in detail. However, given AS cites it as providing the underpinning of its own explanations, it will be important to understand its major elements and to see how far they accommodate sociology's standard practice (remember, AS professes to be about organising what we currently do. If Woodward's account of cause cannot be so accommodated, adopting his principles will inevitably mean re-directing the discipline.) and how far they can be accommodated alongside AS' other theoretical principles.

Using Woodward's own explication as our guide, we take the following to be the major principles of his account:

1 *Scope*: although he adopts what he thinks is a broad definition of the term, not all explanations are causal in form. Explanations you might give of how to drive to Lancaster, the meaning of 'oxymoron' or the reasons why you decided to abstain from voting at the last election are not causal as far as he is concerned. The latter point is very important, because within sociology the attempt to assimilate reasons (for example, justifications or excuses) to causes has had a long and turbulent history. In as much as AS does insist on reasons being treated as causes (see below), it is stretching Woodward's notion of cause beyond permissible use.

2 *Manipulability*: causes are to be conceived not as properties but as variables, that is as having values. If x is the cause of y, then we need to be able to say how y will change if we increase or decrease the value of y. Binary relationships, x is either present or absent, are not fully explanatory in Woodward's sense. The relationship between x and y can be deterministic (y always changes with a change in x) or non-deterministic (the occurrence of a change in y for a change in x conforms to some probability distribution). It is in manipulating the value of x in respect of y (or of x while holding u, v, w constant if these are other contributory causes) that we demonstrate the causal relationship between x and y counterfactually. This demonstration has to be *repeatable and reproducible*.

3 *Intervention*: whilst it is not necessary for an intervention to be possible for the causal account based on it to be explanatory (without the friendly intervention of the Vogons, it is hard to imagine how we might manipulate the asteroid and mass extinction case), the intervention must be conceivable. Such intervention should take the form of a change in the value of the proposed cause and not a change in the cause itself. There is, then, an important notion of "tuning" a variable, and thus the concept of cause can only be applied to variable that can be so tuned.

4 *Invariance*: the relationship between x and y should be (relatively) stable under a range of specifiable conditions. The greater the range of (test) conditions in which this stability or invariance is exhibited the greater the autonomy of the causal explanation. The qualification of relative stability allows for exceptions to the generalisability of the causal relation, allowing us to distinguish between strong and weak causal accounts according to the degree of their autonomy. The quantification of invariance allows us to distinguish the ways in which, though invariant, x has differential effects on y and the extent to which u, v, and w change their causal effects relative to y.

How well do the above principles fit with what usually goes on in sociology and with AS' own principles? To help us get a view of these questions we will re-look at the explanation of a run on the bank provided by Robert Merton, an explanation that has an iconic status in AS.

This is the example in Merton's own words.

> *It is the year 1932. The Last National Bank is a flourishing institution. A large part of its resources is liquid without being watered. Cartwright Millingville has ample reason to be proud of the banking institution over which he presides. Until Black Wednesday. As he enters his bank, he notices that business is unusually brisk. A little odd, that, since the men at the A.M.O.K.s teel plant and the K.O.M.A. mattress f actory are not usually paid until Saturday. Yet here are two dozen men, obviously from the factories, queued up in front of the tellers' cages. As he turns into his private office, the president muses rather compassionately:"Hope they haven't been laid off in midweek. They should be in the shop at this hour." But speculations o f this sort have never made for a thriving bank, and Millingville turns to the pile of documents upon his desk. His precise signature is affixed to fewer than a score of papers when he is disturbed by the absence o something familiar and the intrusion o something alien. The low discreet hum of bank business has given way to a strange and annoying stridency of many voices. A situation has been defined as real. And that is the beginning of what ends as Black Wednesday-the last Wednesday, it might be noted, of the Last National Bank.*

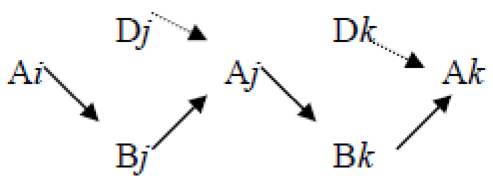
> *Cartwright Millingville had never heard of the Thomas theorem. But he had no difficulty in recognizing its workings. He knew that, despite the comparative liquidity of the bank's assets, a rumor of insolvency, once believed by enough depositors, would result in the insolvency of the bank. And by the close of Black Wednesday-and Blacker Thursday- when the long lines of anxious depositors, each frantically seeking to salvage his own, grew to longer lines of even more anxious depositors, it turned out that he was right.*

> *The stable financial structure of the bank had depended upon one set of definitions of the situation: belief in the validity of the interlocking system of economic promises men live by. Once depositors h ad defined the situation otherwise, once they questioned the possibility of having these promises fulfilled, the consequences of this unreal definition was real enough. (Merton1948 pp194-5)*

The explanation Merton gives is clear enough. A change in the beliefs that depositors had about the financial stability of the bank combined with their own desire to protect their personal interests provided them with sufficient reason to come to the bank in the middle of a weekday and demand to withdraw their deposits. When enough of them had done this, they had created the very situation they feared, namely the insolvency of the bank. We can readily sketch an intervention (e.g. change the depositors propensity to believe rumours or increase the bank's access to short term inter-bank lending) which would produce a counterfactual (the run on the bank would not begin or the demand would peter out as a large number of people were seen to be able to access their savings). At first sight, then, the example fits with the scope, manipulability and intervention criteria. But, if we look more closely, things appear not to be so certain. What motivates the behaviour of the depositors is their beliefs about the bank and their fear of its consequences. For AS, psychological events such as the holding of beliefs and fears

> *.......can be said to cause an action in the sense of providing reasons for the action. A particular combination of desires and beliefs constitutes a "compelling reason" for performing an action. They have a motivational force that allows us to understand, and in this respect to explain the action (Hedstrom 2008 p 326)*

However, understandable though they might be as 'compelling reasons', such reasons are not causes in Woodward's sense. Rather, as we have already seen, they are the characterisations offered when we judge, defend or otherwise evaluate an action. To say that someone had reasons to act in the way that they did is not to say that they were caused to do so, at least as far as Woodward is concerned. To see why this must be the case, we need to go beyond the causal sketch which Merton provides. We could do this using the digraph notation that Hedstrom takes over from Woodward.



The digraph traces how the actions of actor i effect the beliefs of actor j which along with the desires of actor j effect j's actions which in turn effect the beliefs of k which along with k's desires effect k's actions. But what do the directional flows associated with the arrows actually mean? How do one person's beliefs or actions bring about change in another's? Is this process the same as that by which the asteroid brought about the mass extinction, or that which brings about the acceleration of the block on an inclined plane? Can we trace the path of connections at the individual level? Indeed, are there any connections in that sense? Or it rather that we rationalise the behaviour of others and justify our own behaviour by attributing the influence which the actions of actors have on the beliefs and desires of others and how that produces changes in behaviour, including our own? Of course, in ordinary life we do talk of our actions being caused by this or that, but in doing so we are not using the term in the same way it is used in science. This does not make it wrong or inadequate. It is just that cause is used in many different ways in our language, and these are just two of them.

Even if we disregard this argument and insist it is 'in principle' possible to trace the individual connections that bring about the growth of a rumour and a run on a bank, the complexity of tracing such connections makes the task infeasible. But, even if we are willing to accede to arguments such as Donald Davidson's (2001) and assimilate reasons and causes, it is clear that Woodward does not and in fact explicitly rules these types of explanations out from his definition of cause.[11]

What about manipulability, intervention and invariance? To what extent are the contributory causes, the beliefs about the bank, the fears of possible consequences and the desire to protect one's own interests really *variables*? To what extent can we can we conceive of scaling them (other, perhaps, than in an ordinal sense) and tuning them? How would we set about rating beliefs, fears and desires at an individual and collective level and how would we envisage tuning them? This is not a question of ethics; that is, it is not about whether we should intervene to adjust these values. It is a conceptual issue. What does the operationalisation of beliefs, fears and desires look like? The answer, of course, is that it looks like behavioural psychology. This raises two questions. First, is AS content that its causal explanations should in fact be reducible to psychological explanations? Second, is what behavioural psychology measures as levels of belief, desire, fear what Merton means when he talks about those phenomena?

Invariance is the requirement which proves fatal to AS' use of Woodward as a defence of causal explanations. It is invariance and the autonomy of the explanation with which it is associated which determines whether any particular causal explanation is a good one. You don't have to be a follower of Actor Network Theory searching for the exhaustive list of *agencements* to recognise that, as constructed by both Merton and AS, the self fulfilling prophecy is not actually that full (or good) an explanation of the run on the bank. Alongside the beliefs, desires and actions of the people who queue outside the bank and withdraw their money, the conditions over which the explanation has to be stable include the regulations that control the debt to asset ratios of banks, the loans to savings policies they adopt, the consequences of arrangements for inter-bank lending and the problems of cash management, to name just a few. Given the complexity of any social situation, do we know how to separate those conditions which do make a difference from those which don't?[12] Moreover, we cannot resort to the (problem avoiding) device of *ceteris paribus* because we don't

[11] As usual, we need to be careful here. Davidson does not treat all reasons as causes. Your beliefs my cause me to change mine and I might therefore change my intentions. But my intentions do not cause my action.

[12] In passing, we will just note that the end of his discussion of invariance, Woodward discusses Michael Oakeshott's assertion that in history everything is connected to everything else in a reciprocal web of interdependence. As a consequence even if we imagine that the

know where to draw the line between those conditions which are material and those which are not. Merton's bank did not collapse just because a lot of people believed it was in trouble, had their money invested in it and so went to get their money out. Nor was it simply because a rumour started and so created a panic run on the bank. Though both of these are true. A whole raft non-social 'causes' were involved as well.

Finally, the Merton story is set up to be a token causal account, that is a causal explanation of a single event. For Woodward, token causal explanations should explain why the event happened here at this time and not there at that time. Do we know enough, can we find out enough, to enable us to say why the run occurred here and not there in any way that we could call "causally deep"? AS accedes the limited generalisability of causal mechanisms by describing their effects as only "regularly" produced (Arthur Stinchcombe refers to mechanisms as "sometime true" bits of theory). Given this uncertainty over the degree of stability and invariance of any explanation, how do we explain why rumours about a bank's financial state sometimes produce a run on the bank, and sometimes not.[13] How do we know how good the explanation is *this time*? Given that it does not and cannot answer these questions, in the end Merton's causal explanation turns out to be simply a sketch of a type causal explanation. And in that regard, how much explanatory force does it have?

Explanatory force is a property of arguments. And different arguments have different modes of explanatory force. Causal arguments assemble sets of empirically related conditions and their outcomes. As we have seen, the explanatory force is in the regularity of the connection between them. Other kinds of explanations take different forms as, for example, with the practical syllogism where terms are connected by their rationality or intelligibility based upon particular beliefs or norms. In his paper 'Resisting the Force of Argument', Jonathan Adler (2009) quotes Daniel Dennet as follows

> *Surely the following has happened to you, it has happened to me many times: somebody corners me and proceeds to present me with an argument of great persuasiveness, of irresistible logic, step by step by step. I can think of nothing to say against any of the steps. I get to the conclusion, but I don't believe it! This can be a social problem. It is worse than unsatisfying to say: 'Sorry, I don't believe it, but I can't tell you why. I don't know. (Op. Cit. p. 339)*

The point Adler is making is that you have to be predisposed to accept an argument to be convinced by its force. The same is true of an explanation. Unless the explanation answers/addresses the problem you had in mind, then it is not an explanation for you. Its explanatory force depends on what *your* interests are and not on what the interests of the person giving you the explanation are. As Alan Garfinkel (1981)and others have made abundantly clear, what counts as an answer to both why and how questions depends on the context in which either is asked and most importantly on what Putnam (1981) calls the explanatory interests of the inquirer and not the explanatory pre-dispositions of the answerer. That is the point of the oft quoted Willie

French Revolution might have turned out differently, there is no way for the historian to say how history would have been determinately different. In such a world, as Woodward says "there are no causal capacities that remain stable across different contexts" (Woodward p313). And the consequence of that is "...there is no guarantee that every subject matter must be a suitable domain for causal explanation" (Woodward p314). This begs the fundamental question: is the sociology AS wants to develop more like history than physics (or even biology - see below)? Given the wealth of examples drawn from historical studies, we can only infer that it must be.

[13] For instance, In the UK Northern Rock in 2008 and The Co-operative Bank in 2013.

Sutton story.[14] This leads us to ask whether there should be explanatory monism in sociology? Should all sociologists be looking for the same kinds of answers/explanations? Certainly, if there is one truth that is universally acknowledged, it is that sociology exhibits explanatory pluralism and seeking to impose to a monolithic structure is not going to work.

4.2 Causal Mechanisms in Sociology and Life Sciences

Two parallel arguments are offered for the investigation of causal social mechanisms. The first is that it allows sociology to model itself on biology and its affiliated disciplines rather than physics and the mathematical natural sciences. The advantage of this move is that, if successful, sociology could give up what appears to be the hopeless search for explanations conforming to the covering law model of explanation. This would a positive outcome simply because the discipline had proven manifestly unable to discover such laws. However, the attractiveness of biology as an alternative to physics is based solely on the claim by Crick (and repeated by others) that biology seeks to explain phenomena by describing mechanisms. Interestingly, in recent discussions of AS, the modeling of sociological explanations along the lines of biological ones has become much less prominent. We will see in a moment just why this might be.

The second argument is about what might be called the dominant style of sociological theorising; a style which, in the jargon the times, was called "black box theorising" (the term is Raymond Boudon's (1998)). The notion was taken from systems engineering where designers of modular systems often refer to general purpose or limited 'off the shelf' components included in their designs as "black boxes". What they mean is that as long as the component takes inputs in the form which the designed system generates and delivers outputs to the system with the functionality required, there is no need to know or even care about the processes by which these inputs are turned into those outputs. The component can be plugged in and function as a 'black box'. It is not clear that Merton thought along these lines, but certainly Arthur Stinchcombe does. (though neither used the term). For Stinchcombe (1998), general sociological theory consists of rafts of unexplicated black boxes. Observed phenomena such as distributions of educational attainment are explained by variables such as socio-economic status of parents, gender, ethnicity etc with the only linkage between *explanans* and *explanadum* being derived parameters such as a 'causal path' (ordered lists of regression coefficients) or 'explanatory factors' (ordered lists of eigenvectors of a correlation matrix) or a bridging proposition of the 'If we assume the truth of the grand theory we are advocating, it stands to reason that given X, Y would be the outcome'.

These two arguments raise numerous questions. The first, and possible the most pressing, is the extent to which biology and sociology are actually similar enough for the mechanism model of the one to be applicable to the other. Clearly the stronger the similarity, the stronger the case for following biology's strategy. After considering this, we will turn to the matter of black boxes.

[14] For those few who haven't heard it, the story goes like this. A person was trying to reform the then notorious bank robber, Willie Sutton and asked him why he robbed banks. "Because that's where the money is!" was the reply.

A number of things are going on in the search for a connection to biology. First, AS is repeating the age old canard that if sociology wants to be taken seriously as a *bona fide* discipline, it had better model itself on one of the natural sciences. Second, biology has been chosen not because its methods and approach are isomorphic with sociology but because it utilises mechanisms, and since that is what AS insists sociology should do, biology has to be its model. The adoption of biology looks, then, to be a convenience at best, based solely on the assumption the life sciences generally deploy explanatory mechanisms rather than seeking general laws because biologists talk about mechanisms a lot. Some support for this view is given by the fact that commentators such as Bechtel & Abrahamsen (2005) and Machamer (2000) do not say that all science either does or should use mechanisms as explanations (neither do they say that all life sciences do or should), but they do say a majority of life scientists talk as if they do. Moreover, the mechanisms that are described *provide complete or self contained explanations* of the biological phenomena identified. In other words, there is a level of explanation that is self contained and does not require further decomposition into chemical/physical explanation to see how the mechanism's mechanism works. Given the relaxed view of the reduction of sociological explanations to psychological ones of at least some proponents of AS, this is important. In the life sciences, mechanism explanation "bottoms out", to use Machamer's phrase.

In the life sciences, a mechanism explanation is a model composed of three related elements:

- Specified set up and termination conditions. The set up conditions are transformed into the termination conditions by the mechanism. There is, then, a logical and temporal ordering between set up conditions and termination conditions.
- Entities with sets of designated properties. These properties provide the capacities to have the causal effects which the mechanism generates and do so on a regular basis. Entities must be in some appropriate (spatial) relationship to each other (i.e. relative co-location) for the chain of causal effects to work.
- Activities are the processes undertaken by the defined entities which bring about changes in the set up conditions through the causal chains identified.

Two important features follow from this view of mechanism. First, a mechanism will only count as an explanation if it regularly or uniformly produces the transformation. Single case mechanisms or random mechanisms do not count. Second, the causal chain must be traceable right the way through from set up conditions to termination conditions as a series of (physical) entities and processes linked in a causal chain. The account will not count as a causal explanation if any link in the chain is missing.

Proponents recognise that not every mechanism explanation will be a *good* explanation. In her review, Franklin-Hall (unpublished) suggests that such explanations might be prone to three sorts of errors:

- Causation errors: these are mechanisms which appeal to processes which do not explain the changes in causal terms but in other terms such as correlation or association (or indeed other, common causes at a deeper level). Theoretical rigour is the bulwark against causation errors.

- Carving errors: the elements in the explanation (entities and activities) do not have a proper basis in the extant biology but are rather artefacts invented to enable the explanation to work. conceptual rigour is the major assurance against carving errors.
- Zooming errors: the explanation is offered not at the first level reduction of the phenomena being explained but at some other level. Zooming errors can take both micro (eg explanations in terms of quantum entities and processes) or macro forms (eg explanations in terms of structural entities and processes). Empirical reference secures an explanation against zooming errors.

We are not qualified to say if the characterisation given for the Life Sciences is adequate to what goes on there. Neither, just now, are we concerned with it as a philosophically secure general account of what is entailed in explanation. We return to the latter below and there is much debate in the philosophy of science over the former. All we will say is that even though biologists might talk a lot about mechanisms and describe them in their papers, that fact does not demonstrate that instead of searching for general theories, the life sciences describe mechanisms. The theory of evolution and what is known as 'systems theory' are both pretty general and both are widely used in the life sciences (though not perhaps in those areas to which AS looks, namely neurophysiology and genetics).[15]

The key question is whether the specifications set out above are closely enough aligned with how explanations are developed in the sociology for biology to be adopted as a model. Note that this is not the same task (though it is related) as determining whether explanations offered as exemplars by AS (a) conform to the model and (b) are 'good explanations'. For the moment, let us just focus simply on the possible gap between explanation in the life sciences as defined above and the standard forms of explanation in the sociology and thus on what AS would have to achieve to bring about any reasonable level of alignment. The first question, then, is :'Even if we opened up all the black boxes in sociological explanations, would we have explanations which satisfy the requirements needed for explanatory mechanisms in the Life Sciences?' If the answer is negative, this would mean that to use the Life Sciences as a model, Hedstrom and his colleagues will have to do a great deal more than simply regulate the discipline. They will, perforce, have to re-define it, almost from the ground up. Doing that will be a much more deeply radical agenda than has been sketched by Merton and Stinchcombe or avowed by Hedstrom himself.

The question is not about what the sociology should do, but what it actually does. And put in this way, it becomes very easy to answer. Sociology's explanations look nothing like life science mechanisms. There are several reasons for this.

---

[15] Though of course, the whole point of the study of genetics is to show how evolutionary theory actually works!).

- Sociology has no clear, agreed and standardised protocols for determining what are and what are not the relevant set up and termination conditions. Nor, for any set of conditions which are agreed to be relevant, does it have ways of determining, what will count as an adequately complete description of any of them. Different accounts of the 'same' phenomenon begin with different set up conditions and different termination conditions.
- There is no agreement on the ways in which *certeris paribus* assumptions should be allowed to regulate comparability. That is, there is no agreement on how to isolate the phenomena to be included in the explanation. Differently variegated conceptual bundles are deployed on what are held to be 'the same' topics.
- Conceptualisation of individuals, groups, institutions and social structures is highly diverse and while are all held to have effects of one kind or another, there is no agreed description of the capacities of any of these theoretical objects nor how they bring about the effects they are held to have.
- In sociology, the relationships between entities are logical rather than spatio-temporal, and hence, for the most part, the primary form of effect is 'action at a distance'. Entities exert forces on one another more akin to gravity than to the diffusion of molecules or energy. As a result, as we have seen, causal chains of connectivity remain a mystery.
- Temporal ordering is a problem for sociological explanation. To provide such ordering, possible future states of affairs have to be defined as causes of the expectations, anticipations and assumptions of individuals, groups and collectivities so that these, in turn, can act as causes of action. How future events can have current causal consequences is a hotly debated topic in such esoteric areas as quantum mechanics. Certainly, sociology has no argument for how the same approach can be applied in its domain.
- At best, sociological explanations are explanation sketches and routinely fail to trace the piecemeal, step by step, causal path from set up conditions to termination conditions. The reason for this is, of course, that the data required to do so is not available and highly unlikely ever to be. Even supposing we could agree the definition of the relevant causal entities in relation to a run on the bank, how would we obtain the data on all the individuals, group, collective and institutional actions to trace through the move from set up conditions to termination conditions? Even if we can envisage being able to do this for some possible sociological world, is it practicable in the world we have?

As if this wasn't enough, sociological explanation regularly exhibit causation, carving and zooming errors. Explanations invoke principles of statistical association, functional alignment, mutual dependency and structural homology to explain how effects are brought about. Similarly sociological constructs such as ideology, social strata and power are introduced to provide intermediary processes between cause and effect (or to be cause and effect). As a consequence, such constructs are projected onto the social world (another general form construct) and empirical reference achieved by analytical categorisation. The net result is that sociological explanations explain sociology not social life. Finally, the debate over who has or has not committed zooming errors is endemic. All attempts to 'solve' the supposed problem of macro/micro divide have failed. Accounts slide back and forth between the two even when they adopt holism or individualism as a methodology.

To bring about the alignment with biology required for mechanism explanations to be secure, AS will have to undertake little less than a scorched earth campaign. Not just theory, but modes of analysis and forms of method will have to be overhauled even though it is not clear what the acceptable alternatives should be. Without detail on this, the insistence on copying biology would be a leap in the dark based on a prior leap of faith.

4.3 What are Causal Mechanisms?

According to Arthur Stinchcombe mechanisms are "bits of 'sometimes true'[16] theory...or model that represent a causal process, that have some actual or possible empirical support from the larger theory in which it is a mechanism, and that generate increased precision, power, or elegance in larger-scale theories" (Stinchcombe 1998, p267). The qualification on the truth value of mechanisms is important. They are not universal laws but general enough to be useful. The mechanism described in the paper from which we have just quoted is the monopolistic competitive market and we will return to it in a moment. However, because his accounts are intuitively easy to follow and his examples plain, to get a flavour for what mechanisms might be and how they might work, we will start with the doyen of formal mechanism-based explanations, Thomas Schelling.

In *Micromotives and Macrobehaviour*, Schelling (1978) lays out a number of families of (mathematical) models which describe how rational behaviour can lead to unexpected (and sometimes irrational) consequences. The models Schelling uses together with his illustrations are: cyclical functions (how a room thermostat and a moving traffic wave work); critical mass functions (why queues build up at full restaurants, Akerlof's explanation for the predominance of "lemons" in the used car market); and diminishing or increasing marginal return functions (the tragedy of the commons, why no one picks up litter and self fulfilling expectations and conventions).

One example from Schelling's book, the account given of racial segregation in housing, has, like Merton's run on the bank, taken on iconic status in the AS literature. It uses a "near neighbourhood" sub-variant of the critical mass function and is, as Schelling points out, highly generalisable. The only requirements for the function to work are that the variable (race in this case) be dichotomous (ie either/or), exhaustive (universally applicable) and recognisable (publicly identifiable). Alongside these requirements, Schelling introduces two further conditions, one logical, the other stipulative. Within a defined geographical space, both groups cannot be in the majority at the same time. There has to be some distribution ratio. Second, there is

[16] The phrase originates with J.S. Coleman. By it he means that theories are models which are true if and only if the initial conditions under investigation are identical in all respects to the postulates of the model. As Coleman remarks, this is seldom the case.

perfect knowledge of the distribution in the affected population. Two features motivate the model: individuals in both groups have 'tolerance limits' for the proportion of the other type they prefer in their neighbourhood and there is entry and exit from the neighbourhood (i.e. by moving, the population of actors can affect the ratios). When the distribution reaches the tolerance limit of an individual, they will move out of the neighbourhood. The same level of distribution will attract in someone from the majority category. Depending how the tolerance distribution is set, when the model is run iteratively the neighbourhood will reach either an integrated equilibrium or the complete exclusion of one or other category. Even though the set up conditions defined the majority of the population with a level of tolerance for the each other, for housing to become highly segregated along racial lines.

Clearly, the outcome is a function of the shape of the preference structure in the context of the requirements and constraints. Vary it, or either of the latter, and the model will have different outcomes. Without a valid empirical basis to any preference structure or to the constraints imposed on the setting, the model is interesting but will turn no 'cogs and wheels' in the real social world. Schelling offers no evidence (not even the anecdotes he offers in support of other types of model) for the existence of the tolerance distributions he uses let alone why such preference structures should be taken to have stronger causal force than, say, employment structures in the neighbourhood, or quality of housing, or any other factor identified in studies of housing patterns. Instead, what Schelling gives are mathematical functions applied to 'social scenarios'. They might provide an interesting way of thinking about alternative mathematical descriptions of rationalised behaviour, but they provide no insight into the actual causal processes affecting racial segregation in actual communities.

Although Merton and Schelling are the most widely cited exponents of social mechanism explanations, it is to Jon Elster that AS turns for an account of 'the mechanics' of the mechanisms which they identify. The "nuts and bolts" of social behaviour as Elster calls them in his book of that title (Elster 1989) and in its later revision *Explaining Social Behavior* (2007), are the causal devices by which the mechanisms work.[17] These devices are the beliefs, desires and opportunities of individuals. They provide the causal motivations for human action and so explain social behaviour. It is important to recognise at the start that it is Elster's *arguments* about the causal force of beliefs, desires and opportunities which AS relies on. At no point does Elster provide a detailed demonstration of how a specific social phenomenon or process actually occurred and what the relevant causal beliefs, desires and opportunities for that case were. In other words, he offers no demonstration of the efficacy of his "tools". Rather, he relies on toy stories, summarised versions of historical and sociological examples. By being informal, Elster hopes to make his arguments clear and simple though he recognises that he runs the risk of looking lightweight. His defence is that he is pointing the way forward not undertaking the journey himself. This is, of course, the usual social theorist's ploy; state the programme of work needed but leave the (hard) work of carrying it out to someone else.

[17] In examining Elster's arguments, we will concentrate on the later work. It is, as he makes clear, a revised and expanded version of the earlier one.

Here are some examples of the kinds of social "puzzles" or problems which Elster wants to explain.

- Why are people reluctant to acknowledge, to themselves and others, that they are envious?
- Why did the French victory in the 1998 soccer World Cup generate so much joy in the country, and why did the fact that the French team did not qualify beyond the opening rounds in 2002 cause so much despondency?
- Why is sibling incest so rare, given the temptations and opportunities?
- Why do passengers tip taxi drivers and customers tip waiters even when visiting a foreign city to which they do not expect to return?

Clearly these are all social in some sense. Whether they are sociological problems and hence in need of sociological explanations we will leave to one side. Suffice it to say that Elster's aim is to show how the requisite combinations of beliefs, desires and opportunities provide the causal mechanisms which explain them.

Elster's argument has a number of components. We will take each in turn and examine how well it might provide the support which AS needs. We will begin with the principles around which the whole construction is designed.

1. Explanations of social behaviour have to be couched in terms of the actions which individual persons can take. For Elster, this is 'methodological individualism'. By this he is not saying that groups, collectivities and social formations do not exist. That would be ontological individualism. He is simply saying that whenever we describe or explain what they do, we are using summary shorthand for what in fact individuals, usually in concert, do. Elster doesn't argue for methodological individualism (except by dismissing non-individualistic explanations as non-explanatory or weak). He assumes it. For him methodological individualism makes sense because it is correlated with one of his other principles, reductionism. The two have a symmetry and, on some interpretations, are logically connected. In that sense, what Elster is committed to is a strong methodological individualism.
   Given its commitment to Coleman's 'explanatory boat' of causes and supervenience, strong methodological individualism could be a somewhat difficult principle for AS to adopt. Strong methodological individualism rules out any *explanatory* appeal back from individuals and their actions to collectivities. But this is just what the boat is supposed to provide. However, for AS to drop strong methodological individualism would raise questions about the robustness of the beliefs, desires and opportunities triad.

2. Elster espouses strong reductionism. By this he means that, in principle, social explanations are re-writeable in psychological terms which in turn are, in turn, in principle re-writeable in biological terms, and so on all the way down to physics. Ultimately, all explanations will be

reducible to explanations by physics. For him, it does not matter that we do not have many (or indeed pretty much any) examples of how this reduction is to be achieved. He does discuss some examples of how biology might provide explanations of patterns of behaviour, but they are largely speculative and are certainly not uncontestable. Explanations have to operate at the level of individual actions (that is what methodological individualism means) and hence psychology and biology "must have a fundamental importance in explaining social behavior" (2007 p36). Elster's first level reduction poses AS very few, if any, difficulties. Individuals have psychological predispositions (a social psychology, if you like) which explains what they do. AS simply ignores Elster's insistence that the principle is iterative. In fact, AS has to ignore it otherwise *ab initio* it would be cutting the ground from under its own feet. As far as Elster is concerned, some day sociology will cease to be an explanatory discipline (as will biology, chemistry etc). All they will be is descriptive; physics will provide all the explanations we will need and can have. Until that point, of course, sociology, psychology and biology (to name just three) have temporary explanatory status. This is not a position either sociology or AS are likely to be willing to endorse.

3. Explanations of action are causal. That is, to explain a phenomenon w is to provide the practical reasoning through which a relevant list of antecedent phenomena, a, b, c which in juxtaposition produce w. The connection between a, b, c and w is one of *logical implication*. That is, given a, b, c we have no choice but to accept that w follows. When applied to sociology, this is expressed in the following kind of practical reasoning. Given the psychological pre-conditions a, b, c, it is rational to do w. That is, 'do w' is logically implied by the propositions a, b, c. Elster does not say if the conditions a, b, c are simply necessary or sufficient or whether they should be necessary and sufficient despite the fact that in philosophy, at least, how these two relate to causal consequences has been earnestly debated. All he says about the relationship between w and a, b, c is that the latter constitute a 'mechanism' for producing w and "if this kind of thing happens, here is the kind of mechanism that might explain it" as well as "if this mechanism operates, here is the kind of thing it can produce" (p1). Both of which are, at best, a pretty loose explanations.
   It follows from this principle that any explanation which cites consequences cannot be explanatory (that is, by definition, cannot be explanatory). Consequences are not causes. This is the reason Elster rejects functional explanations (at least in sociology. He does let them creep into biology on the grounds that they are couched in terms of feedback mechanisms and thus dynamic processes of cause and effect).
   As we have already indicated, the causes that Elster has in mind for sociological explanations are specific congeries of beliefs, desires and opportunities held by individual actors. He is also willing to accept intentions as causes, but as we saw earlier that simply raises questions regarding the temporal ordering of cause and effect. Of course, what needs to be explained is not that psychological conditions a, b, c, cause w but how they actually induce/produce that action. That is what mechanisms are for.

4. Explanation is singular. The upshot of principles 1 - 3 is that there is only one kind of explanation, and paradigmatically it is that provided by physics. This is a rejection of what earlier we called the interest relativity of explanation. Interest relativity proposes that explanation is a social practice and that what will count as an explanation of w depends on the context in which it is asked and given and most importantly on the interests of those who ask for and those who give explanations. Different sets of conditions will work as explanations in different circumstances and for different purposes.

   More extensively, the singularity of disciplinary explanation implies a constancy hypothesis which says that the phenomenon which is described and explained by sociology is *analytically identical to* the phenomenon of the same name as described and explained by psychology, biology etc etc. This matters for AS since it has to be able to offer an account of why it is relaxed about the reduction to (social) psychology (why the constancy hypothesis holds there) and not about the further reduction to biology and beyond, if, as it seems clear it must, it refuses to accept that step. That refusal is implied in the adoption of the model of explanation from biology along with the associated principle that its explanations 'bottom out'.[18]

Given these principles, what exactly is an explanation? An explanation is a description of a set of conditions which acting together either serially or in concert produce the phenomenon in hand. Such conditions form a causal chain. The existence of a phenomenon or event is explained if and only if we can specify the causal chain which brought it about. Elster suggests that most sociological explanations are actually about facts not phenomena. Fact w is brought about by facts a, b, c. The rate of take up of higher education in the UK among different social groups, say, is explained by listing rates of disposable income, rates of literacy, rates of pre-schooling, shapes of earnings curves of the occupational positions available and so on. To explain the fact about higher education take up what has to be given is the description of the causal linkages, and this will involve a description of the grounds which individuals have for the decisions which they make; i.e. the beliefs and desires they have and the opportunities they perceive.

What, on Elster's view, would not count as a proper or satisfactory causal explanation? Explanations which do not provide the causal linkages (obviously). Explanations in terms of correlations (but this is just to reiterate the old saw 'correlation is not causation'). Statements which assert determinacy, that is, the individual had no real choice in the decision that was made to carry out the action in hand. This is, of course, interesting since compulsion is central to causal explanation in their home scientific domain namely the analysis of material objects. Social action cannot be explained by the kind of compulsion which the push-pull mechanisms of physics and engineering have. But if not, what does this mean for Elster's reductionist strategy? Elster also asserts that answers to "Why" questions are not causal (this despite the fact that nearly all the puzzles he starts with are couched as "why" questions!). However, a little reflection shows "why" questions often do prompt causal explanations of the kind that Elster seeks. The last group of putative explanations

[18] See Philip Gorski (2013) for a different way of construing this issue.

which Elster rules out are predictions. Predictions are not themselves explanations (who thought they were?) but can be based on explanations.

Having justified his view of causality by presenting a series of broadly philosophical arguments, Elster turns to investigative methodology and hence to the practicalities of undertaking actual sociological investigations. He recommends a six step strategy to isolate social causes.

1. Determine if the phenomenon at issue is in fact the case. Since most social phenomena are 'general facts', this means we have to ascertain their factual status.

2. Choose a theory (set of interrelated causal propositions) which looks like the best bet as an explanation.

3. Specify a hypothesis that links the *explanans* to the *explanadum* (y to x) so that the x *follows logically* from y.

4. Build some counter cases which are similarly logically connected.

5. For each alternative, try to *refute* it by pointing to testable implications which are *not* observed. This gives lateral support to the hypothesis.

6. Strengthen the proposed hypothesis by pointing to additional testable implications that are observed (novel facts). If these implications relate to other next level down questions, then the hypothesis has excess explanatory power. Elster calls this "support from below". The theory can also gain support from above if it can call on other bits of theory that fit with it.

Let's take each of these and see how AS (or any kind of sociology for that matter) might stand in regard to them.

The first task is to determine the factual status of 'the factoid' that is to be explained. The immediate problem we face is that all the methods we might use to determine 'factuality' are themselves the methods used to produce the facticity of the factoid in the first place. And, as endless studies of sociological methods have shown, the facticity of official statistics, social survey results, the outputs of case analysis and so on are socially constructed in and through the methods used to produce them. We cannot disentangle the results from the methods (neither can physics and all the other natural sciences). Nor can we recover those methods from the research reports wherein the results are published. That Elster knows this is demonstrated by the following acid comment on sociological research strategy.

> *Once a scholar has identified a suitable mathematical function or a suitable set of dependent or independent variables, she can begin to look for a causal story to provide an intuition to back the findings. When she writes up the results for publication, the sequence is often reversed. She will state that she started with a causal theory; then looked for the most plausible way of transforming it into a formal hypothesis; and then found it confirmed by the data. This is bogus science.*

> *In the natural sciences there is no need for the ''logic of justification'' to match or reflect ''the logic of discovery.'' Once a hypothesis is stated in its final form, its genesis is irrelevant. What matters are its downstream consequences, not its upstream origins. This is so because the hypothesis can be tested on an indefinite number of observations over and above those that inspired the scholar to think of it in the first place. In the social sciences (and in the humanities), most explanations use a finite data set. Because procedures of data collection often are nonstandardized, scholars may not be able to test their hypotheses against new data. And if procedures are standardized, the data may fail to reflect a changing reality. (Elster 2007 p.49).*

None of this means that sociology is impossible. It is just as possible as physics, chemistry, biology and the like. It is simply that the metaphysical realism that Elster presumes as the underpinning for his principles is on very unsafe ground. If we are to determine how the facts are independent of the ways we determine what the facts are, we have an impossible task. Given AS' endorsement of realism (though the variant seems to shift from occasion to occasion), Elster's first step is likely to be a major challenge; or at least it will be if the aim is to put sociology on the same (rigorous) footing it is assumed the physical sciences are on. Of course, one could always withdraw the requirement for realism. But if AS were to do that, what then would be the basis of the argument for an equivalence with the natural sciences?

The tripwire contained in the second step is not actually to do with the causal propositions. It is "the best bet" requirement. To have a "best bet", we have to have a field of runners and riders to bet on. That is, we have to have an array of equally plausible explanations to put in a preference order. Listing implausible explanations would be both self-defeating and attempting to rig the outcome. One can imagine any number of alternative "scenarios" which might, in ordinary life, be used to provide plausible explanations (in fact many of the mechanisms Elster invokes are just such scenario forming devices). Relationships between John and Jane look strained. perhaps they have had a tiff; or there is a problem with one of the children; or they are having money troubles; or..... And we might, perhaps though this is a stretch, imagine constructing equally plausible alternative causal scenarios in physics. What is certain is that in sociology we have trouble enough finding just one plausible account to act as a causal explanation, never mind constructing an array of equally plausible ones. Getting sufficient data to make one explanation stand up is continually found to be too great a challenge. It turns out the 'logic in use' or 'logic of discovery' adopted by sociology (as is implied by the quotation we have just cited) is not one of opening up and then narrowing down options. It is, rather, one of getting just enough to data to correspond with the theory that we are assuming is the case so that we can declare confirmation has been achieved.

Step 3 looks straightforward but, as we all know, looks can be deceiving. The task is to build an account where the *explanadum* logically follows from the *explanans*. How are we to determine that some statement w logically follows for some other statement a? What are the standards against which we are to make this judgment and how firm are they? Notice this is about securing the relationship between statements not securing the relationship between social facts or social events. Mapping the relationship between the statement and its target (for want of a better word) is an entirely different thing as we will see in our discussion of ABM. The point, of course, is that the interpretation of logical implication and logical compulsion is itself a matter of convention. Logical implication may be enshrined in a set of conventions or rules defined in one or other predicate calculus, but seeing the force of the implication, the requirement to accept the implication, is a matter of interpretation, as Lewis Carroll's story *The Tortoise and Achilles* makes abundantly clear. Setting out the logical implication of w on a is a matter of explication. Where does this stop? At what point is enough enough? As practical researchers (and it is the practicality of Elster's proposals we are now discussing) we will always be able to do so. But on what grounds independent of our own (subjective as Elster would term it) judgment will that decision stand? The practice of sociology provides a normative framework for making such determinations but such relativity appear at odds with AS' conception of objectivity.

Step 3 is hard, if not impossible. Step 4 wants us to do it several times more! The only way this will work is through the relaxation of the requirements of rigour specified in the programmatic statements about realism, objectivity and scientific method which have either been explicitly adopted or implicitly endorsed by AS. The trouble is that AS will then look just like every other sociology endeavour it is attempting to replace. It too will be riddled with black boxing and hand waving.

Steps 5 and 6 are not about constructing explanations but securing or justifying them. What is important to note here is that Elster now changes the basis of the justification or grounding of the causal connection. In steps 1 - 4, the linkage is one of logical implication. What he wants now is empirical verification. What grounds or secures logical implication are the rules of logical or rational inference which govern truth preservation across propositions. As Achilles put it, "If you accept a, b and c, then you *must* accept d". Empirical verification is about existential status. Conditions a, b and c might well be in place but that does not guarantee that d will necessarily be in place even though it is logically implied. This is what opens up the space for Elster to say that causal explanations are non-determinate (and hence depend on choices). However, the question is not about non-determinacy but whether we can argue for the evidential grounding of logical implication, or whether asking for that grounding is actually a category error. Even if it is not, as we will see, all sorts of methodological and practical matters come crowding in to render such a step extremely risky.

Step 5 is about weeding out the alternatives to the main hypothesis by means of counterfactuals. If alternative p was correct then we would see events/processes q, r, s. We don't see q, r and s therefore p is not correct. There are two problems here. First, since the connection we are testing is a logical one, we are being asked to prove is a negative. But that is something that borders on a logical impossibility. Of course, Elster isn't actually saying that. He wants us to find evidence (or a lack of it) for q, r, s. The test is not a logical one but an empirical one. This then throws us back on the ability of sociology to specify (all) the conditions relevant for r, s, and t such that p would follow from them. But sociology has no protocols which would exert such control. Even in psychology (which does heavily use experimental protocols), the investigative techniques required to assure us that we had ruled out any and all conditions that might be operating to prevent p in the presence of r, s, t are not in place. In the end, all AS could do to implement Elster's step 5 would be to adopt some canon of materiality as a methodological principle. But just what would this be? And how would it be established and secured? Moreover, would it not leave AS (and Elster) open to the same order of criticism that they use about levels of significance and confidence in statistical analyses? The problems with step 6 are the inverse of step 5. How will we know that there is not some antecedent condition f that is producing the empirical phenomenon that is the valid target for the novel implications? Clearly we can't. The net result is that the logical connection remains underdetermined.

It seems that following Elster's programme for developing and securing causal explanations is likely to lead sociology, and not just AS, into a practical and methodological quagmire. Perhaps if we look at the use of mechanisms in such causal explanation, we might find a way of avoiding getting bogged down?

The first thing to say is that what Elster means by a mechanism is very different to what AS seems to mean and is certainly nothing at all like mechanisms as they are described in biology. For Elster, they are (roughly) "frequently occurring and easily recognisable causal patterns that are triggered under generally unknown conditions or with indeterminate consequences" (Op. Cit. p.36). Notice the two adverbial clauses. They are triggered under (generally) unknown conditions and have indeterminate consequences. As we have seen, a mechanism-based explanation in biology requires specification of all the relevant initial conditions and the consequences must follow from those conditions. if AS wants to use biology as its shield for the use of mechanism explanations then it cannot (or so it seems) invoke Elster's version of mechanisms as the form such explanations will take,

For Elster, the indeterminacy of outcome preserves the centrality of choice and absence of compulsion. Things didn't have to turn out the way they did. But since they do have to turn out some way and they have to do so through causal chains, we cannot (again *ex cathedra*) know all the conditions under which they might operate in any specific case. The kind of thing that Elster means by a mechanism is childhood socialisation. Barney has a difficult personality and cannot form friendships because his parents spoiled him. Charlie is an alcoholic because his parents were. Belinda is bright and does well at school because her parents are academics. And so on. Childhood socialisation (we learn from our parents whose own actions re-enforce that learning) is the link between what the parents did and what the child does. But, of course, Barney, Charlie and Belinda might not have turned out the way they did. We all know families where the children are very different from their parents (and sometimes that is a good thing too!). But, and this is the central point, since we cannot set out all the conditions under which action is taken and we cannot trace through the complete set of connections which do or do not bring about a consequence, we don't know why Georgia did not do well at school whereas Belinda did even though they both had the same parents (or, at least, we don't, if we are using childhood socialisation as the causal explanation). Something else must be at work, but we don't know what! What kind of explanation is that?

One way of resolving this might be to appeal to a further set of causal factors, say innate talent. Belinda has it but Georgia doesn't. We now have two causal forces at work, nature and nurture. Our problem now is that while we can say (along with commonsense) that nature and nurture must both have *something* to do with how Belinda and Georgia turned out, we are completely unable to say just how much each had to do with it. And, moreover, we are (quite rightly) legally and professionally forbidden to undertake the studies that might tell us. So, we might be able to list the two, three or four causal forces at work, but what we can't do is say whether they are additive, multiplicative, operate in a ratchet fashion or whatever and what each actually contributes. We can produce lists, but are such lists really explanations in the sense that AS wishes them to be?

The unpacking of causal mechanism such as childhood socialisation involves articulating the psychological states/processes/phenomena which account for our individual choices and hence actions. Elster lists the main kinds: motivations, self interest or altruism, myopia and foresight, beliefs and emotions. All these can "induce" (to use his term) the desire to take particular actions. Thus they are the mechanisms. And, of course, we recognise them and use them ourselves to explain other's behaviour. Georgia is governed by immediate gratification (we might pompously say) while Belinda is not and so Georgia does not see the long term value of education whereas Belinda does. The same goes for accounts in terms of motivation, self interest (the infamous "Well he would wouldn't he?" explanation of Mandy Rice Davies), beliefs and emotions. However, to explain a specific pattern of actions we have to nominate particular instances of the types; which motivations, which emotions, which particular aspects of self- interest and so on were held by which individuals? And accessing these *as independent data* to support empirical verification rather than as either attributions or self justifications seems more than a little far off. Without such independence, we are thrown back on the same rationalisations which form the basis for our commonsense accounts of action. Our explanations will be interpretations and no more. Elster is quite comfortable with this since he includes interpretations in his account of explanation. However, what this does to AS' search for a robust 'scientific' basis for sociology is an another matter entirely.

The psychological well springs of action are modulated through two filters. These are the final component in Elster's scheme. The first is filter is desire and the second opportunity. I might have the psychological reasons to act in a certain way (be an alcoholic, be a difficult person, be good at school) but do the opportunities present themselves to be so and do I want to take advantage of them? Once again in identifying desires and opportunities, we are in the world of first person descriptions and are faced with all the challenges we have just discussed. Remember, all this is in service of weeding out the alternative plausible explanations by reference to evidence for their existence. We can attribute desires and opportunities but how do we know the first were actually in place and the second were actually perceived? But even with desire and opportunity, the social context may prevent the action being realised or force it to be channelled in specific ways. This is because we do not live as social isolates. Our actions take place in a field of action undertaken by ourselves and others. Trust, norms, and organisations shape what we do by shaping the decisions we make. We have 'internal' desires and opportunities and 'external' constraints and facilitators. For Elster, to explain social behaviour is systematically to set out the psychological causes of the desire to act in a certain way and the perception of an opportunity so to do *and* the social causes which act to facilitate or constrain us in doing so.

If we could do this, what would such explanations look like? None of the cases, stories, anecdotes or historical events cited by Elster are analysed in anything like the depth required for us to use them as exemplars. The puzzles he started with surface here and there in his discussion and return at the end, this time packed up as behavioural puzzles and their explanation. Here is how they turn out.

> *Why are people reluctant to acknowledge, to themselves and others, that they are envious? Answer: because they care about their self-image and because envy, in most societies, is near the bottom of the normative hierarchy of motivations.*
>
> *Why is shame more important than guilt in some cultures? Answer: because a society that has not conceptualized guilt will also display less guilt behavior.*
>
> *Why did the French victory in the 1998 soccer World Cup generate so much joy in the country, and why did the fact that the French team did not qualify beyond the opening rounds in 2002 cause so much despondency? Answer: because surprise is a magnifier of both positive and negative emotions.*
>
> *Why, in Shakespeare's play, does Hamlet delay taking revenge until the last act? Answer: because Hamlet is subject to weakness of will and because the tension could not be resolved before the end of the play.*
>
> *Why is sibling incest so rare, given the temptations and opportunities? Answer: because natural selection has favored a mechanism inhibiting sexual desire for same-age members of the opposite sex in the same household.*
>
> *Why do military commanders sometimes burn their bridges (or their ships)? Answer: because they expect that their opponent, knowing that they will be unable to retreat, will abstain from a costly fight.*
>
> *Why do people often attach great importance to intrinsically insignificant matters of etiquette? Answer: because they think that someone who deviates from the norms does not care what they think about him.*
>
> *Why do passengers tip taxi drivers and customers tip waiters even when visiting a foreign city to which they do not expect to return? Answer: because the thought that others think badly about them is painful.(Op. Cit p.453)*

Each of the "because" statements does contain a mechanism of sorts, but it is hard to see these as the rigorous social scientific explanations that AS wishes for. If they are explanations, they are not very good ones.

4.4 AS and Causal Mechanisms

So far we have been discussing the authorities which AS cites for the use of causal mechanisms in explanations. But what of AS itself? How does it view mechanisms and is this view symmetric, consonant or even broadly similar to those of the authorities cited. In an extended discussion, Hedstrom and Ylikoski (2010), summarise their position in the following way. Notice how the list of characteristics offered is interestingly different to the definitions offered for mechanisms either in the Life Sciences or by Jon Elster.

- A mechanism is defined by its consequences, the effects or phenomena it produces. The characterisation of which effects/phenomena are thought to be produced by what mechanism is then vital.
- Mechanisms are causal. The entities have causal effects. However, such effects might be probabilistic. Any causal account will not be exhaustively descriptive. Irrelevant details will have been abstracted away in order to focus on the core elements. Such abstraction is achieved by counterfactual analysis. If some component makes no difference to the effect to be explained, it can be ignored.
- Mechanisms have a structure: they are composed of entities, properties, processes.
- There is a decomposable hierarchy to explanatory mechanisms. Explanatory mechanisms in one disciplinary area (say sociology) rest upon mechanisms in other disciplinary areas (psychological and biological sciences, say) which in turn rest upon mechanisms in the physical sciences. This does not imply an infinite regress (for H&Y at least) since any discipline's mechanistic explanations "bottom out" on mechanisms it is not its role to explain. All that is required is that such mechanisms "really exist" (p.52). To explain what they mean by "really exist", they fall back on a version of Roy Bhaskar's "critical realism".
- The aim of a mechanism is to explain a set of 'facts'. These may be empirically gathered (as with Stinchcombe's example) or stylised (as with Schelling). The preference is always for explanation to be in terms of empirical facts but this remains a challenge for the sociology and so it may be necessary to develop the array of mechanisms needed with sylised facts in order to motivate the sociology research needed to evidence them. However, without accumulation of empirical evidence about the entities, properties and processes the account given remains at the level of mechanism-based story telling.

In a separate paper (Hedstrom 2008) Hedstrom reinforces the layered or nested structure of mechanisms and explanations. Although we can identify structural effects and the mechanisms which produce them, nested within these mechanisms are mechanisms that explain the actions of individuals (we will return to the structure/action linkage below). He then summarises the programme and its virtues as follows.

> *.........we must explicate the mechanisms that explain the actions of individuals, and which are nested within these "structural" mechanisms. These types of action-related mechanisms may also be characterized in terms of their entities (and their properties) and the ways in which the entities are linked to one another. The core entities are different, however, and now include entities such as beliefs, desires, and opportunities of the actors. But the explanatory logic is the same: we explain an observed phenomenon, in this case an individual action, by referring to the mechanism (that is, the constellation of beliefs, desires, opportunities, etc) by which such actions are regularly brought about.*

> *Why is it so important that we identify the mechanisms that appear to generate the outcomes we observe, whether they be the actions of individuals or collective outcomes that result from the collective or sum-total of the actions of numerous individuals? For one thing, identifying the details of the mechanisms tend to produce explanations that are more precise and intelligible. In other words, we can only really understand and explain what we observe by referring to the mechanisms involved.*
>
> *Another important reason is that focusing on mechanisms tends to reduce theoretical fragmentation. For example, we may imagine numerous theories (of voting, social movements, etc.) that are all based on the same set of mechanisms of action and interaction. By focusing on mechanisms we may avoid unnecessary proliferation of theoretical concepts. We may also be able to place in relief structural similarities between processes that at first glance seem completely dissimilar.*
>
> *Finally, an understanding of the mechanism involved in an outcome is what permits us to conclude that we are dealing with a genuine causal relationship and not simply a correlation. As Glennan (1996:65) has emphasized, "two events are causally connected when and only when there is a mechanism connecting them." Without the ability to identify such a mechanism we cannot conclude with any certainty that an observed regularity is indicative of a genuine causal relationship. (Op. Cit. p323)*

In summary, then, mechanistic explanations are nested series of causal accounts within which mechanisms made up of entities such as beliefs, desires, perceived opportunities, emotions etc cause individuals to undertake courses of action. Such mechanisms take forms like self fulfilling prophecy, preferential selection, rational imitation and monopolistic competition which then produce patterns of socio-structural effects. In contrast to the Life Sciences we have no detailed specification of set up and termination conditions and no detailed tracing through of the detailed deterministic processes which connect them. In contrast to Elster, the reductionism is half-hearted and social (and hence collective) phenomena are attributed causal properties. In contrast to both, the requirement for empirical validity of the causal mechanism can be relaxed and replaced with stylised depictions of the facts. Given such disparities, without a lot of strong argument it seems hard to accept the claim that the search for causal mechanisms which AS is promoting can be legitimated by reference to the rigorous example of the Life Sciences and the (social) philosophy of Jon Elster.

4.5 Examples of causal mechanisms

Of course, all of these reservations might evaporate if the descriptions of causal mechanisms provided by actual studies undertaken in the name of AS were sufficiently convincing. We will look at two which are widely cited as exemplars of mechanism explanations: Gould's (1993) study of trade cohesion and militancy in the Paris Commune and Bearman and colleagues' (2004) study of romantic relationships in a High School. Both centre on the effects of norms on producing collective behaviour. We will also briefly review some of the studies offered in the *Handbook of Analytic Sociology* edited by Hedstrom and Bearman (2009).

4.5.1 Chains of Affection

This study was carried out as part of a wider investigation of the spread of sexually transmitted disease among adolescents. 573 adolescents at a High School were questioned about their recent (ie in the previous 18 months) romantic and non-romantic sexual relationships and asked to identify their partners. The dominant pattern of relationships which emerged took the form of a *spanning tree network*; that is, a network with a clearly identifiable central spinal ring with short branches. The researchers compared it to the network architecture of a rural telephone line. The puzzle was how to explain this pattern. What 'social rules' might the students be following which could produce such a distinctive pattern?

The researchers simulated a number of rules to see if they might produce the observed network. For example, partner choice might be random or might be based on a preference for particular attributes which both partners share (i.e. homophily). Neither replicated the pattern. A third simulation based on homophily was run in which there was partner sharing (2x2 sharing) giving a 4 link cycle. This did generate a network somewhat similar to the observed one. Finally, simulations were run on a model with links ≤ 3. This 3 link model was expressed as a rule that (from the male perspective) that boys should not have sexual relationships with their prior girlfriend's current boyfriend's prior girl friend. Although individuals expressly chose their partners on the basis of perceived attractive personal attributes, the pattern produced by these choices conformed to operation of the above social rule.

Bearman *et al* rationalise this rule by suggesting that peer group status is important to the students and hence its loss avoided. Two forces are at work; homophily which is a preference for partners who are similar to oneself and hence one's immediate peer group and the avoidance of what the researchers call "seconds" (i.e. recent partners of an individual who is closely linked through the network). The mechanism producing the pattern (the social rule) rationalises the unarticulated (and possibly unarticulatable) preferences of the students. *If* they were following a norm such this *then* they would produce this pattern. The rationalisation is not implausible and does make the pattern intelligible. However, and this is key, it is not a norm recognised by the adolescents in question. As a consequence, although the rules describes the data very well, its external or empirical validity as the motivation for *their* behaviour must be deemed speculative at best.

4.5.2 Trade Cohesion: the mechanism that wasn't.

Accounts of the patterns of insurgency during the Paris Commune of 1871 face a puzzle. Conventional explanations point to the effect of craft union solidarity as the cause. Because craft unions were tightly knit groups organised to preserve their exclusive rights, these seem the mostly likely social formations to have been militantly committed to the Commune. The puzzle is that, for some period prior to the Commune, the strength of these bonds had been loosening with social identity increasingly being found in terms of generalised relationships such as a shared class position. If the specific ties of the craft were being replaced by the diffuse ones of shared class consciousness, how could trade cohesion explain militancy?

This is the puzzle that Gould sets out to solve. His conclusion is that trade cohesion can't explain militancy. Instead, relying on his own earlier work (Gould 1991) he proposes that it is neighbourhood-based networks of social ties that were at the heart of the insurgency. The evidence for this (negative) conclusion is drawn from the analysis of an 1740 court dossiers of the trials of members of the Paris National Guard which led the insurgency and who were found guilty. Using a number of variables reflecting relevant parameters of craft membership, working conditions and residential neighbourhood, Gould undertakes a multiple regression analysis. Here is his conclusion:

> *These results send a strong and, from the point of view of the orthodox perspective on artisanal activism, troubling message. If the artisanal activism thesis were correct with respect to working-class involvement in urban insurrection, we would have observed that the most active trades were those with the greatest residential cohesion and lowest ratios of workers to employers. Instead, the data demonstrate the contrary: craft-group organizational capacity, as measured through these two variables, was negatively related to participation in the 1871 insurrection.(Gould 1993 p 746)*

In this earlier study, Gould looked at the residential base of the Paris National Guard. Although members were recruited on a residential basis, there was a great deal of "shuffling" of cohorts during the early weeks of the Commune. Again, Gould's method is a regression analysis of network links among the residential areas (*arrondissements)* from which Guards were drawn. Once again, here is his conclusion:

> *These findings show that insurgents in different neighborhoods influenced each other's degree of commitment to the insurrection through the network of links created by overlapping enlistments. High levels of commitment in one area enhanced commitment elsewhere when enlistment patterns provided a conduit for communication and interaction (Gould 1991p 726)*

What is important for our purposes is that nowhere does Gould provide data on the beliefs, desires, opportunities and constraints which members of the Paris National Guard might have held (nor those of members of a craft union either). What he presents are tables of regression coefficients which he analyses in the usual way. As an example of explanation through the provision of causal mechanisms, Gould's work looks to be more like what AS calls black box theorising than anything else! This does not mean it is weak as an example of the kind of analysis social history could aspire to; far from it! But what it does mean is that it is hardly an exemplar of causal mechanism explanation.

4.5.3 Mechanisms in the Handbook of Analytical Sociology

It is not unreasonable to think that AS would take the opportunity of presenting a full volume summation of its work so far (Hedstrom and Bearman 2009) to provide strong and detailed examples of how its account of causal mechanisms and the micro-macro link might be cashed out. Social movements, orchestras, campaigns, social media and even open-source software development could all be described as collective action which is the outcome of individuals actions but which cannot be explained in terms of the individuals alone. Certainly a number of chapters in the *Handbook* begin as if that is exactly what they *are* going to provide; detailed explanations of how specific instances of collective action arose from the action of particular individuals. At least one, we might hope, ought to provide a case which traces through all the detail of the proposed mechanisms to show how they were generated by the actual beliefs, desires and opportunities of a set of specific individuals. Alas that is just what does not happen.

Delia Baldassarri (Baldasseri 2009) thinks collective action is made possible though "...the co-occurrence of individuals interest and group identity, by, first producing a shared representation of the collective good, and, second, indicating a consistent course of action." Co-occurrence of interest and identity looks like a mechanism in AS' terms, though Baldassarri doesn't call it such. Having specified it as her phenomenon, co-occurrence disappears from her account to be replaced by a discussion of formal models, in particular game theory, and only reappears several pages later as a conclusion to a discussion of the free rider problem. Such overlap between individual and collective interest makes collective action possible as the by-product of the emergence of collective identities from patterns of social action. Any consistent course of (collective) action ultimately depends on the availability of a shared representation of the collective good and only the presence of a collective identity can generate confidence that the individual will be capable of fulfilling their own interest in the long as well as the short run.

Co-occurrence of individual and collective interest looks like a mechanism. It is structured like a mechanism and it has the supervenience modality of a mechanism, so the title of the next section, "The origin of shared identities and collective interests", raises our hopes. Moreover the section starts on just the right note: a specific case, street protests in Berlin in 1989. But that is the last we see of it or any actual example. Instead, what we are given is discussion of the assumptions built into (computational) 'models' of the interplay of the micro and the collective, all of which seem to start from the truism that people will prefer to be with people who have opinions much the same as their own. Groups coalesce from this selectivity not by the blind binding of individuals but from social interaction, persuasion and separation. Whatever else she does in summarising studies of collective action, Baldasserri sheds absolutely no light on questions of mechanism action, causality and supervenience. While there is lots of reference to studies of collective action, whether (a) these are species of AS and (b) the accounts they give are any more that informal causal story telling is impossible to determine.

James Moody (Moody 2009) has even less interest in presenting the outline, sketch or even gesture of a real case. What he discusses is a catalogue of graph theoretic models of dynamic networks and their configurations. Two 'examples' are examined; an invented story about cocktail party talk and information flows and the (urban) mythical 'small world phenomenon'. Both are used to illustrate representational formalisms and their properties not as a putative description of actual data (ie real cocktail party information flows or real attempts to reach Barack Obama, Pope Francis and the postman in Kakudu National Park in 6 steps each).

Given sociology's obsession with power and status orderings, one would think dominance hierarchies would be a prime topic the display of AS' distinctive mode of analysis. Certainly Chase and Lindquist (2009) are keen to assert how general and thus important formal and informal hierarchies are and, indeed, how their preferred social model of linear structured hierarchy is an improvement on all previous models in both behavioural biology and sociology. Their model is the interaction process model. Basically, this model, which was developed in studies of chickens, suggests that hierarchies emerge through dyadic and triadic transitive and intransitive dominance relationships and not simply through interactions based on individual attributes. Unfortunately, we are offered no studies of social organisations (other than chickens) to demonstrate the applicability of this model/mechanism. The final paragraph of their paper neatly summarises their contribution.

> *The essence of what we have done in this chapter is to point out that there are fundamentally two different theories for the forces that produce social organisation in small groups and anmals.......If the second theory is a more adequate view of social organisation, it has fundamental implications for how we might study the formations of hierarchies, networks and other kinds of structure..... (Op. Cit. p 586 - 7).*

Unfortunately, without the data and methods which they go on to say are needed, we have no way of determining if the theory they offer is more adequate as a description of actual cases rather than a preferable formalism.

The same story holds for the other studies in the *Handbook* that purport to be about casual mechanisms and the macro-micro link. None have any obvious interest in specifying mechanisms, demonstrating causal entities and supervenient properties. In fact, only Field & Grofman (2009) spend any time discussing an actual study of a social group but that is to demonstrate how sub-groups protect the individual from cross pressures to conform to different social norms (about parenting). Basically, their finding is that people don't know or care what most other people other than their friends did and took care to manage their own behaviour so that those who were not their friends didn't know what they did. They provide interesting ethnographic insights as to how such parents managed this. However, the provision of causal mechanisms as set out in AS' programmatic statements, it certainly is not.

4.6 Taking Stock

What are we to make of all this? From the discussions we have reviewed, three main conclusions seem to press themselves on us:

1. Despite its protestations that it intends to re-frame sociology around a strategy of developing middle range theories from the studies of particular detailed cases of individual action, few if any new studies are on offer. Instead, when actual cases are discussed, the descriptions given are of generalised types engaged in different courses of action. Where beliefs, desires and opportunities are offered as explanations, they are summarised as the typical beliefs, desires and actions that typical actors are assumed to hold. Since explanation and description through typification is the standard mode of sociological analysis, it would seem AS is offering nothing new. Sociologists past

and present have all provided accounts of social activity cast in terms of the beliefs, desires and opportunities (among other things) of members of the societies they studied. Even a paper with the really promising title of 'Theoretical mechanisms and the empirical study of social processes' (van den Berg 1998) turns out to be no more than the usual AS critique of structural models and other forms of quantitative analysis with no actual examples of real empirical investigations. In the end, AS exhibits not so much methodological individualism as a strategy but abstracted individualism; the same defect for which it berates the rest of sociology.

2. Of course, what is new (or asserted to be new) is the proposal that such beliefs, desires and opportunities are causes in the same way that springs and catches, spindles and latches are and that self fulfilling prophecies and the like are mechanisms in the same sense that locks and mousetraps are. However, when we look at what is required to sustain those comparisons, we find that sociology can neither satisfy the conditions for causal descriptions in the physical and material sciences nor those in the biological sciences. This is not because the analytic process of simplification have not been developed sufficiently, but, as Woodward points out, that sociology has not developed the modes of analysis required to describe the complexity of social life in the detail necessary for simplification not to be distortion. It seems, then, that when AS talks about causes, it does so in ways that are somewhat remote from either the physical science or the biological ones.

3. The central ambition of AS is to find a way of reconciling the grand polarisations of macro and micro theory and analysis in sociology. The key terms in its reconciliation are causal decomposition and supervenience. This duality is designed to account for how collective life emerges from individual life. What AS fails to see is that both polarities are defined as social *a priori*. We cannot conceive of an isolated individual single-handedly possessing or inventing a social institution such as language or a normative system. While the ontogenesis of social life might be an interesting ethological and perhaps archaeological question, it is not a sociological one. Sociology begins with social individuals acting within the institutional frameworks which define any society.

At this point, It would seem our conclusions must be pretty bleak. The central concepts of AS are muddled and its claims to affinity with any of the natural sciences weak. Moreover, the studies that it undertakes and celebrates turn out to be no more and no less conventional than the sociology it attacks. There is a sense that AS understands this, which it is why it has turned to agent based modelling as a new investigative tool to ground its approach. Before we look to see if agent based modelling actually can help AS overcome the difficulties it faces, we want to step back and ask about the character of and requirements for analytic comparisons of this kind. To our mind, it is a failure to understand these requirements as much as anything else which has brought AS to the state it is in. Without considerable clarification of what is involved in analytic comparison, we fear adding agent-based models to the mix is only going to make things worse.

5.0 MECHANISMS, MAPPINGS AND METAPHORS

Unlike 'syllogism' or 'supervenience', concepts like 'cause' and 'mechanism' are not simply philosophers' terms of art. The varieties of uses we put them to are deeply embedded in our normal, everyday ways of understanding and speaking about the world around us. Choosing one to be paradigmatic at the start will only lead to interpretive difficulties later. 'Cause' and 'mechanism' are particularly susceptible, if not to misuse then certainly to over-stretched use, when particular interpretations are given such paradigm status in academic disciplines. This is not simply because they are used in variously different ways in our ordinary usage. Underlying that usage is a pattern of conceptual relationships which even what Philip Gorski (2009) calls the "philosophical" usages draw upon. As we have already suggested, it is devices such as the spring and catch of the mousetrap or the spindle and latch of the door lock that act as what Eleanor Rorsch (1977) called prototypes for that pattern. This device prototype is what has been translated and refined by the physical and mechanical sciences.

When, we ordinarily use the concepts of cause and mechanism to describe events outside the physical, material and mechanical world, we are using a metaphorical mapping. 'The slamming of the door made the baby cry' we say and mean no more than just as routinely a releasing the spring on a mousetrap causes the trap to close so routinely loud noises cause babies to cry. The actions follow one another. Such metaphorical use is part and parcel of our usual way of describing, explaining and predicting and, as such, is mostly unproblematic. However, when we build and extend upon such ordinary metaphorical use to provide analytic descriptions and explanations, it is important to be clear just what is being claimed. Without such reflection, all too often we can mistake metaphorical description for what might be called a 'literal' one. When that happens, the comparative essence of the 'dead' metaphor drops out and we talk not *as if* the phenomenon had the properties described in the metaphor but seem to assert that it does.

To prevent this kind of category error, when we are offering an analytical account which trades upon commonsense usage, it is important to work out first:

1. What are we trying to say about the phenomena under investigation. In what ways are we trying to make them intelligible (and for whom)?

2. Whether the description we give really does make the phenomena intelligible in the way we and our interlocutors want.

These two considerations point to Alan Garfinkel's (1981) conception of the interest-relativity of explanation which we mentioned earlier. We don't have to reach for examples as arcane as the Zande's view of what will answer the question 'why did the hut fall down?' (termites or witchcraft?) or Willie Sutton. Explaining why the child is crying by saying the door slammed is a perfectly good and full explanation. So is saying that there was a run on the bank because people panicked. Both are perfectly good commonsense explanations. Equally good commonsense descriptions are suggestions such as the conservatism of 'revolutionary' parties is an inevitable consequence of political machinations and causes the need to maintain party discipline (the commonsense version of the Iron Law of Oligarchy?). It is only if we believe there should be a single non-metaphorical explanatory form for all disciplines that we feel our analytic uses of terms like 'cause' should be identical (or nearly identical) and require that the mechanisms and causes we describe be somehow just the same as those used elsewhere in the physical or biological disciplines.

Recognising and accepting the interest-relativity of explanation allows us to be much more relaxed about the myriad ways in cause and mechanism are used in sociology and with the many different kinds of explanations deployed there (see Reiss 2009) . Trying to bring all these uses all under a single rubric has all the hallmarks of earlier ventures in theoretical regulation. The stipulation of a singular descriptive form, it will be remembered, was one of the central myths/dogmas of Positivism and empiricism.

Once we see that talking of social mechanisms is a mapping of this kind, we can ask what we mean by the description. Just what elements from the prototype are being carried over in that comparison? The point is that tropes provide only a limited congruence between what are otherwise quite dissimilar things. In saying that a colleague has a razor sharp mind, we are obviously not saying that you can shave with it or that you could cut your fingers on it. We are saying something about our colleague's incisiveness in teasing apart complex problems. The comparison is meant to illuminate or even appraise only some of our colleague's faculties. The use of the metaphor carries no explanatory or other epistemological weight. If AS is using the metaphor of 'mechanism' as a limited trope in this way, then it is largely harmless. Just as we can *describe* or make intelligible what our colleague is good at by comparing his or her analytic skill to that of the sharpness of a razor, so we can describe a pattern of social actions as being like a lock, a mousetrap or the gearing of a waterwheel. If this is all that Elster and AS mean by "nuts and bolts", "cogs and wheels", we can have little objection. An inept comparison it might be, but everyone are is free to choose their own modes of description.

However, if we want to say that the notion of a social mechanism is *more* than simply a broad descriptive image, and that self fulfilling prophecy, rational imitation, relative deprivation and so on are the formal analogues of mousetraps and locks, then we have to say exactly which characteristics are being carried over in our analytic version from the root concept of mechanism to the analogue and how they are being mapped onto the phenomenon in hand. But at this point the link ceases to be a simple comparison and becomes one of formalisable symmetry where the mapping can be made in both directions. The challenge then is to say how relative deprivation is formally symmetric with a ratchet or pulley when we say it has the effect of 'amplifying' social unrest? Exactly what relationship is being proposed (isomorphism? identity?) and with regard to which properties? Are we saying that we expect to be able to derive equations for the operation of relative deprivation in the same way we can for the operation of ratchets and pulleys? Or is it simply that they all provide ways that effects can be magnified. But, of course, that is not an analytic description, let alone an explanation.

As we saw in our discussion of James Woodward, once we start to talk about symmetry, we have to address the question of invariance. What is being held invariant across the symmetry relation? And how is that property bundle being translated in that mapping? The thing that is core to the idea of mechanism and is invariant across ordinary use is the suggestion that causal devices are designed. Mechanisms are purposefully designed to have particular effects. A log which accidentally gets jammed between the ground and a boulder teetering on the edge of a cliff is not a brake mechanism; a wedge pressed against a car wheel to stop it from rolling is. Purposeful design is central to the notion of mechanism.

The thread of purposeful design implicit in the meaning of mechanism helps us understand why the notion can so easily be deployed in biology. Knowing biology's most general theory, we can all fill in the missing design component. Random variation and natural selection produce biological phenomena *as if by design*. Our neural processes can be said to be as much the product of random mutation and natural selection as the shapes of the beaks of Galapagos finches. Given the centrality of 'design' or 'as if by design' to the intelligibility of the mechanism metaphor, the stress which AS places on the importance of causal powers, emergence, supervenience and so on, entirely misses the point. It is the purposefulness of design which makes the comparison work. As with the traditional 'hidden hand' account of markets in economics, outcomes are produced *as if* by design.

AS wants social causes to be more than metaphors. They are the analogues of biological and physical explanations. But just how is a run on a bank analogous to a lock or a mouse trap? Although, despite Sica's (2004) arguments, both are equally real, the nature their effects are quite different (this is the ontological issue underpinning the mapping). Self fulfilling prophecies are not constructed as if by design. A self fulfilling prophecy is a social category; its existence turns upon it being used as an explanation or description. Clocks, locks and mousetraps exist independently of our putting them to use in our explanations. Moreover, clocks, locks and mousetraps work the same way every time we (re-)use them. Self fulfilling prophecies don't. The concept of self fulfilling prophecy is deployed only at a very high order of generalisation (one step below the most general - social action). It is the categorial equivalent is 'tool' or 'instrument' or 'equipment' or some other general category, not mechanism and certainly not mouse trap. The question we are belabouring here is simply 'At what level and for what purpose is the comparison being made?'. With a mechanism, if set up conditions x, y, z are realised and then 'click' the mechanism works (given lots of other conditions too). That, it seems, is all that Merton and AS are after - the 'click' effect. Merton wanted such 'devices' because his objective was to use sociology as an instrument of policy change (remember it all started with racial segregation). He was arguing for the need explicitly to change the institutional structures which sustain prejudice - that is, for a clear change in public policy. In the rhetoric of policy promotion, nothing has much stronger plausibility than the notion of a hidden mechanism.

If, like the Life Sciences, we want to say the force of the analogy is the description that some effect is produced as if by design and if we want to say that the mechanisms are causally real, then we also have to say just what the sociological analogues of random mutation and natural selection are. It is only when we have identified these processes that we can have the debates over realism, emergence, supervenience, causal efficacy, causal depth, reduction and so on. The challenge is to show in what respects an unfolding of collective preference for racial balance in a neighbourhood, or the pursuit of self interest, or propensity to shape our actions to what others are doing is analogous to random mutation and natural selection. Despite all the rhetorical flourish with which AS appeals to the example of biology and the Life Sciences for its use of mechanism, no guidance is given on this.

The knotty problem of the basis on which comparisons are made, and what is involved in the translations between comparators, what is being compared with what and for what purposes, is at the heart of AS's programme. Both a self fulfilling prophecy and a mousetrap are causal mechanisms, but what do they have in common? How can the properties of the one be rigorously translated into the properties of the other? It is at the heart of AS in another way too. Agent Based Modeling (ABM) has been picked out as the one research modality in sociology which can provide AS with the rigorous mechanism-based explanations of social action that it seeks. ABM seeks to translate informal, discursive descriptions of social action into formal, symbolically structures ones. How can these translations be made to work, again without distortion or oversimplification? If the methodology of AS is critically dependent on ABM, as it seems the major figures in AS now think, how does ABM ensure the mapping between the informal and the formal, the analogy and the analogised? In the next section, we take up these questions.

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.

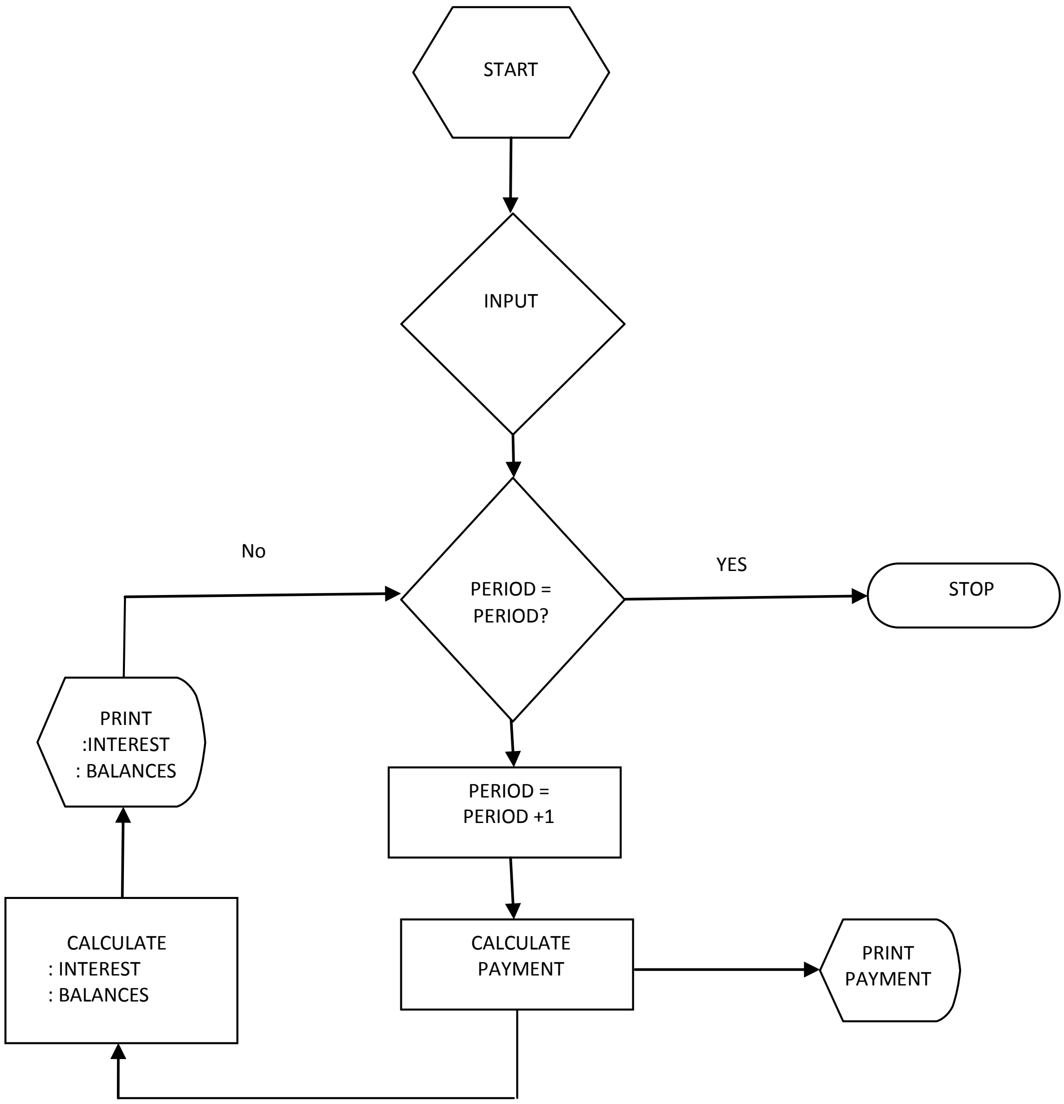


Figure 1 Flow Chart for Loan

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

```
Loan-working.py - C:\Python notes & exercises\Loan\Loan-working.py
File Edit Format Run Options Windows Help

# 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)

Ln: 54 Col: 0
```

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.

Figure 3

```
Loan-working-maths-stripped.py - C:/Python notes & exercises/Loan/Loan-working-maths-stripped....
File Edit Format Run Options Windows Help

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

Ln: 45 Col: 23
```

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.

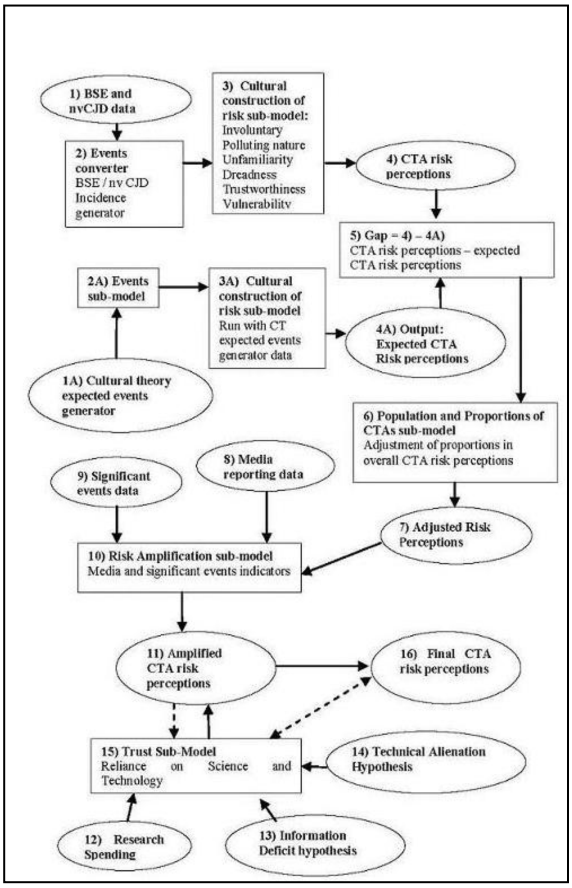
1. Related sets of external events generate specific information about particular possible risks.

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.

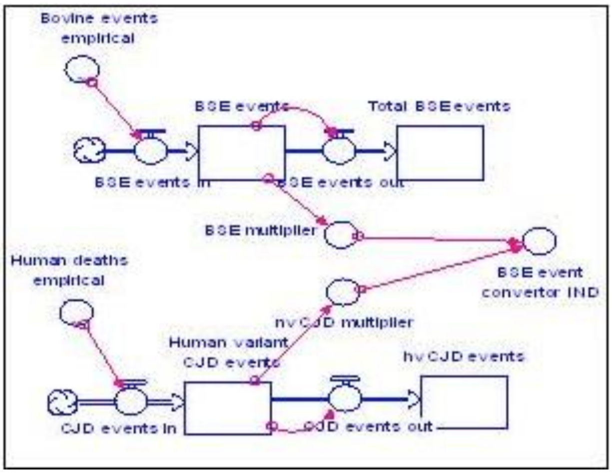
2. The rates of occurrence of these events are the inflows of each type of event.

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:

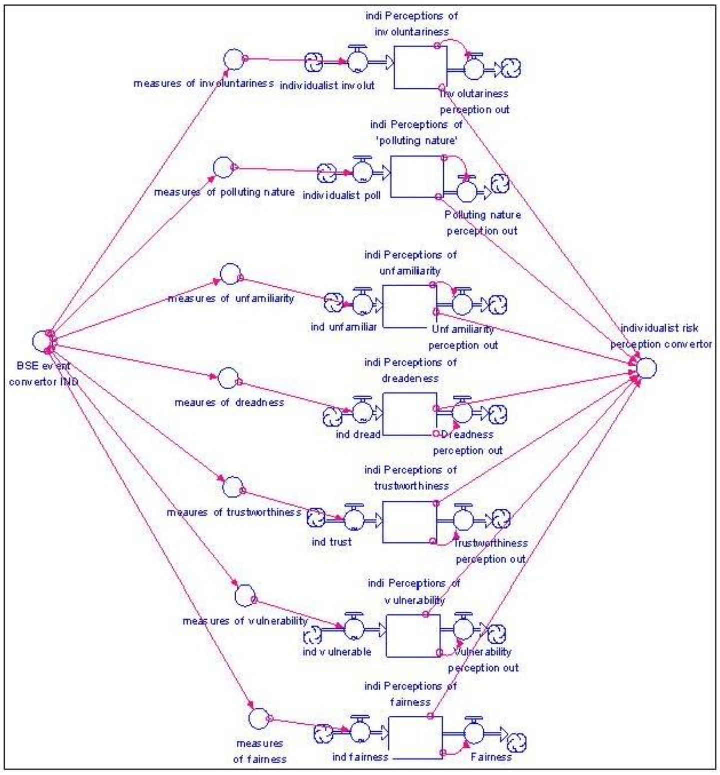
[27] What is interesting here is that (inadvertently?) Bleda and Shackley treat the BSE panic an inductive evidence for the validity of the variables they extract from the literature. Using Polya's terminology, this then becomes a buttress for the credibility of the use of the variables in the BSE case.

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.

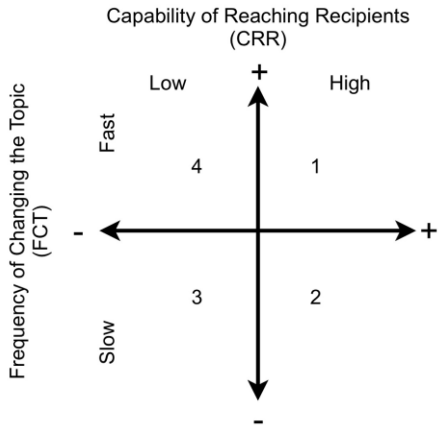


Figure 7 Paradigmatic Strategies

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.

PART D: SWANS AND DUCKLINGS OR LEOPARDS AND SPOTS?

11.1 Crisis as a way of life

Like many similar efforts which have preceded it, AS is convinced that despite all the contentiousness and disarray, it is possible to facilitate a unification of sociology. Ugly duckling it might be but beneath the feathers all stubby and brown, there is a swan waiting to emerge. We are far less sanguine. It seems to us that if sociology looks like a duck, quacks like a duck and walks like a duck, then it probably is a duck. Or, to change the metaphor, the problem with transforming sociology is in getting the leopard to change its spots. We don't know if sociology can or cannot become a science. What we do know is that it doesn't really seem to want to. Embedded in the culture of the discipline are traits which systematically subvert any such proposal for transformation into being a science.

A quarter of a century ago Richard Schweder and Donald Fiske (1986) thought that sociology was in an 'uneasy' state which might best be described as a crisis. The main reason was its failure to deliver the universal generalisations framed in the deductive nomological structures which were then thought to be the hallmark of scientific explanations. This failure was taken either as proof of the persistent immaturity of the discipline or a good reason to give up entirely on the idea that sociology might be a science; a conclusion many have concurred with and acted on. Today, once again the question of whether sociology is in crisis is being posed, though this time it is motivated less by epistemic worries than by the fashionable conviction that sociology is a critical discipline and, as such, it deserves a significant social and political role. Ironically, this view prevails at a time when sociology actually seems to be increasingly marginalised because of supposed threats from (a) its own dispersal into enclaves within adjacent and not so adjacent disciplines and (b) non-sociological organisations and institutions which are gathering of 'big data' on social life.

We mention Scheweder and Fiske not because they used the image of crisis but to underscore a more important point. The questions which preoccupy AS are perennial ones and what AS represents is not the uncovering of a novel general problem in sociology's methodology but just the latest version of a well established *strand* in sociological thought, namely that the discipline should adopt the form of a mature (usually natural) science. For our part, like many others we think that while sociology has several deep problems, that it is not sufficiently like a science is *not* one of them. it would be absurd to deny we are often quite critical of contemporary sociology, but this is not because it falls short of being a proper science but because it lacks good arguments for the claims it makes about its findings and their implications.

We recognise though that AS and others have a different view and no doubt would want to debate with us. Indeed, one of the reasons we took an interest in what AS was proposing was because we recognise we might be wrong about the prospects for a scientific sociology and wanted to see if we were being proved wrong. As is apparent from our discussion, thus far, at least, we don't think so. AS was of interest for another reason. We are sympathetic to any form of sociological thinking that recommends a dose of simplification as a way of disentangling the Gordian knots that the discipline counts as its problems. However, we would expect the pay-off from this simplification to be quite modest and often deflationary. We believe it will be a great deal harder to make sociology into a science than some sociologists seem to think. In that respect, it is our conviction that AS is mostly engaged in misplaced effort. It is only setting out a plan for a scientific sociology, not producing it.

The one lesson from Thomas Kuhn that hardly anyone ever mentions is that mature sciences aren't developed from blueprints. The exemplars that become the focal point for an integrated complex of normal science just appear within a field and then displace the preoccupations which previously dominated. Further, simply because sociology resembles the early history of some now well established normal sciences doesn't mean it will automatically develop its own transformative exemplar. For that to happen, someone has got to do the work of delivering the exemplar. For all we know, this might well be taking place. However, past experience tells us it is far more usual for sociologists to tell us what needs to happen and then refrain doing it themselves. Their manifestos typically and misleadingly offer 'science: the easy way' while at the same time proving just how difficult it is to improve much, if at all, on the current situation.

When Schweder and Fiske were writing, by and large the generalisations sociology was trying to force into the deductive nomological format complied with Dibble's first law: namely, some do; some don't.[31] Many manuals were produced offering templates for how to form deductive nomological structures but rather than setting sociology off on a new footing, all they tended to do was induce scepticism as to whether the theories so constructed did anything more than superficially mimic the required models, and even if they did do anything more, whether the inferences they supplied were either interesting or useful.

11.2 Never mind the quality, look at the volume!

The one insight we think can be taken from all the talk of crisis in sociology is the observation that the discipline's empirical research does not seem to be cumulative. There are multitudes of studies and projects but their results do not consolidate into a unified corpus of knowledge. The point of the remedies for theoretical reformulation such deductive nomological models ormiddle range theory, is to find ways of connecting research findings together through the medium of an integrating theory or model. Unfortunately,

[31] in 1964 Berelson and Steiner produced an 'inventory of findings' from the behavioural sciences, extracting some 1045 'findings about human behaviour we thought had some decent claim to substantiation', presumably only a small number relative to an already voluminous collection of studies. Berelson thought that the 1045 could be reduced to "a 3-proposition grand summary: (1) some do, some don't; (2) the differences aren't very great; and (3) it's more complicated than that".

deductive nomological modeling yielded neither generalizations that could be put to use nor the unification of research. And it looks like AS will share the same fate.

However, these failures didn't extinguish the wish for generalisation and unification. AS (inter alia) offers its own way of revitalising the aspiration by dispensing with what it sees as the misguided idea that scientific laws are universal generalisations. Instead, it insists respected science uses mechanisms which, in their 'sometimes true' character more closely resemble the kinds of generalisations that sociology commonly comes up with .[32] A not-unsympathetic Amazon customer complained in his review of the *Handbook of Analytical Sociology* that the statement of AS' core principles is a statement of methodological principles only. Explanatory principles are notably absent. it seems the problem that AS faces is the same one that confronted the enthusiasts for deductive nomological models – providing a rationale for the adoption of the form is the easy part. The hard questions bear on what is going to be go into the format. In the end, sociology's reformers often turn out to be precisely that, re-formers; their main effort being the recasting of existing sociological materials – ideas and studies - within the form they prefer while at the same time leaving it hard to determine what added value is actually gained. The logic is one that says since existing sociological materials *can* be recast, they should be rather than demonstrating that there are sociological findings and other materials which are crying out to be to be analysed within that form, ideas which *just can't* be set out in any other way. A number of times, we have mentioned the way in which AS recurrently appeals to a small number of precedents such as Schelling's models. Such analytical lash-ups serve more as proof of concept demonstration of the enhanced explanatory power of the AS approach. Just as is the case with many of sociology's other supposed 'paradigms', these precedents are less Kuhnian exemplars than interesting one-offs.

Like everyone else, AS wants to improve sociology's explanatory power, but, again like everyone else, it treats explanation as more a matter of applying an explanatory form rather than making intelligible something that has previously defied understanding. Sociology doesn't really have problems that defy understanding – it is as prolific as it is because it doesn't run up against phenomena which are just so hard to figure out such that they require the adoption of some new kind of explanatory machinery. AS, after all, makes its fundamental explanatory principle the BDO scheme which is pretty close to a default mode for everyday explanations of people's actions, namely that people respond to situations on the basis of their reasons. To take another example, Arthur Stinchcombe is a prominent point of reference for AS and has been a strong advocate of more stringent forms of theory and yet when he tries to explain why there is always resistance to organisational innovation, his argument is that innovations unsettle existing organisations by threatening some of the things that people in them value such as their status symbols (one obvious example being reduction in the amount and kind of office and work space that is theirs).[33] Stinchcombe's analysis is reasonable and readable and makes some interesting points as it ranges over different aspects of the effect of and problems in changing established relations, but the only unifying explanatory principle given is the banal

[32] Dribble's law is still the model, though!

[33] See Stinchcombe (1986)

idea that "resistance to administrative innovations is mostly due to (more or less) rational anticipation that concrete interests (career prospects, risks of failure, chance for exceptional success, etc.) are endangered " (Op Cit p. 221). The unifying theme of the discussion is the setting out of possible variations on this central motif with the main explanatory work consisting of how instances from organisational life can be recognised as instances of self-protecting resistance. Interesting connections are made and interesting examples given that certainly bring to mind aspects of organisational change one wouldn't have thought of, but these are ones which fit readily with. and to an extent derive from, what we already understand about how organisations work, and certainly fit comfortably within a wide range of existing forms of sociological explanation as well as being heavily dependent on common sense familiarity with organisational life.

Jon Elster is someone else that AS treats as a fellow traveller. Elster, however, tends to arrive at what, for him at any rate, are disappointing conclusions about the prospects of developing effective general theory in sociology. He explicitly admits that his own attempt to understand social solidarity and to generate a general theory of social movements falls short of his ambitions and consequently suggests that his own unsuccessful efforts are as good as can be expected. In this, Elster rather like those who, treating their own theorising efforts with which they have now become disillusioned as if they were somewhat more than the sketchy and simplistic accounts they really are, conclude that society is too complex and motile ever to be captured in a general theory (or 'Grand narrative'). Such self-confessed failure serves less to determine the limits of what sociology might achieve than as an illustration of the widespread tendency of sociologists to bite off vastly more than they can chew (to repeat James Davis' complaint about the general preference for problems that sociologists are in no position to solve that we cited in Part A). Just to be clear, though, we are not suggesting a counsel of hope for a grand narrative, only reservations about any presumption that the production of general theories is both quick and easy.

11.3 Silk purses and sow's ears.

It has been suggested that one reason for the relative weakness of sociology's findings and conclusions compared to those of the sciences is the technologically impoverished nature of the instruments that it uses (Turner and Kim 1999) though, of course, sociology doesn't really use 'instruments' in that sense at all. When sociologists talk of 'research instruments', they usually have in mind some type of social survey which, after all, is just a way of asking people questions. When they do use instruments based upon genuine technologies, these tend to be domesticated versions of the familiar audio and video recorder (even so, in our experience lots of sociologists can be unduly impressed with the results so gleaned). This all leads us to think that even if they were able to deploy the most sophisticated of instruments and technologies, the quality of the results generated would be pretty much the same as now.

It is for this reasons that we suggest the use of computing tools in sociological analysis will not make significant, indeed any, difference. Of course, the computer is still recent enough to count as an impressive technology, and the capacity of computer systems to capture and process 'big data' is currently stimulating great expectations (as well as acute anxieties that sociology might lose out relative to business and public bureaucracies that have the resources to acquire vast quantities of social data and powerful means to analyse them. The computer is also seen as opening up new ways of doing scientific research through the use of simulations rather than experiments and studies. Clearly, ABM, for example, provides a (potentially) powerful way of calculating the cascading and compounding consequences of elementary recursive operations, which fits well with AS' idea of what sociology needs to do, namely to track the accumulating consequences of the causal effects of individual actions. However, the usual problem remains. Powerful scientific tools come with their own requirements for proper and effective use, and it is here sociological repurposing of these technologies runs into trouble. The rationale offered is that these sophisticated devices will force sociologists to rethink and raise their game so as to satisfy the instruments' requirements. However, the usual response is precisely the reverse. The requirements of the technology are relaxed in order to accommodate the sociology being processed. The two examples of ABM modelling which we examined indicate just how this happens. Both aim to connect with sociology through implementation of its theories, and both choose what are complex and, like most sociological theories, vague, if not obscure, about how they might yield specifications of the determinate causal chains connecting the phenomena they address. In both cases what is implemented is a version of the theory which is determined more by the ease with which it can be represented in the simulation than it is by the necessities of capturing the theory it is purporting to explore. Salgado et al's 'chatty game', for example, serves to do little more than remind us of the insightfulness of Wittgenstein's remark about how problem and methods often pass one another by.

11.4 The virtue of aspirational realism

The maxim 'cut your coat to fit your cloth' is unlikely ever to be adopted as sociology's motto. [34] Failures of unrealistically ambitious efforts don't have to be projected on any very great scale to show how difficult it is to shape sociology's conceptual and theoretical questions sufficiently sharply to allow telling and definitive answers. But rather than causing us to reflect on why we have fallen so short, the failures are taken as demonstrating the complexity of both problems and phenomena and as a consequence act as a stimulus to pose yet more questions we can only answer in the loosest possible way.[35]

The difficulty of providing definitive answers involves more than simply making the questions more complicated. There are also deep problems concerning our ability to ensure that the data we can collect captures (or maps onto) the phenomena we are interested in exploring. As we have just suggested, sociology has a limited repertoire of investigative methods and tools. Very often, these cannot capture the phenomena we are interested in which would, in any case, require very complex evidence to be documented. The fact is that the tools – which, as Turner and Kim noted, are mainly variants on talking to people – are used despite their known inadequacies. The persistent slippage between the intended and actual objects of study (just one of the many well known instances of slippage throughout sociology!) is usually taken as something to be lived with unquestioningly. Earlier we recounted how James Davis complained that it is rare for any two

[34] We think Conversation Analysis is a rare exception to this.

[35] We'd be apt to say, adding new questions to the huge number that are already being begged.

sociologists study the same thing, which is doubtless one reason why replication is rare, but the form of much sociological research also serves to make replication impractical. Sociological research involves collecting raw materials, processing that raw material for publication as data, and then making claims made about what the data evidences and its relation to the problem originally addressed by the inquiry. Though 'data collection' occupies a disproportionately large place in sociology's methods manuals and reflections, as well as of the distribution of effort within research projects themselves, there is no reason to suppose that, in practice, data collection is a highly disciplined procedure. As Harold Garfinkel (1967) demonstrated, it is usually impossible to recover either the precise procedures used to assemble the raw materials or the procedures used to transform them into publishable data. The procedures themselves are not normally captured, and are likely to be almost as little known to the data gatherers as they are to their audience. Garfinkel also noted that the data presented in sociology is often such as to accommodate different, even opposed, conclusions regarding what has been done, demonstrated and achieved (especially when more than one of the varied perspectives in sociology are applied to the same data).

If there is much room for – indeed, relatively free - interpretation in the analysis of data, then there are reasons for thinking that there is also ample space for equally free interpretation in both the collection of raw materials and in their conversion into publishable data. On those relatively rare occasions when there are opportunities to compare raw materials and data, the relationships between the two can seem highly problematic and available at least to differences of interpretation[36]. It is not a matter of saying that generally the presented data in sociological studies is seriously at odds with its raw materials. It is simply to say that for the vast majority of sociological studies there is no way of telling from what is published whether it is or is not. Again, this is well known but seen as just another condition to be lived with and, for most disciplinary purposes, disregarded. Problems in sociological research methods are never solved generically and once for all so that subsequent researchers will not have to contend with them. They are repeatedly addressed and *resolved* in the context of the demands and difficulties of an individual project and fixed in ways which meet the needs of that project.

We recognise we are treading a fine line here. While we seem to be agreeing with the complaints of those who want to the discipline more scientific and have enumerated some of the very conditions they wish to reform, we do not repeat them in order to criticise but to indicate how unrealistic the reformers are about the situation they confront and the extent to which what are, for them, regrettable states of affairs are, for a great many of their professional colleagues, unremarkably normal and entirely acceptable features of their practice. There is no point in proclaiming that because they don't prioritise making close and systematic connection with other research studies, sociologists are derelict in their scientific duty if they don't see such demonstration as needed for their purposes or required to evidence their achievements. Indeed, they mostly don't recognise, at least in practice, the demands that a 'more scientific' orientation would impose. Rather than regretting the complexity, obscurity and verbosity of many central sociological contributions, many

[36] In *Reinventing Evidence in Social Inquiry*, Richard Biernacki (2012) aims to make a general point about the untrustworthiness of coding practices applied to cultural phenomena, and in doing so provides a couple of striking and very detailed examples of these problems

sociologists welcome these characteristics, often being able to discern elements of profundity within (one or some of) them, and happy to be challenged to explicate what their favoured sociologies are saying. As Alfred Louch pointed out long ago[37], if, in the mature natural sciences, theory is used to throw light on researches, then in sociology an important function of researches is to make intelligible what the theory is saying.

It is a simple fact that because of the uncertain and often indeterminate relationship between what sociologists say and what they purport to talk about, sociological conceptions are not transitive. The fact that one understands and accepts what (selected) sociologists say does not entail that one knows what the social world is actually like. Peter de Grace and Leslie Stahl (1990) in writing about software development, argued that programming skills are narrow but deep. In contrast, you might say that sociological skills are prevailingly wide but shallow.[38] In terms of Schutz's (1976) typology of the distribution of knowledge, whilst many sociologists like to think of themselves as 'experts'[39] it seems to us that they are much closer to the 'well-informed citizen'[40], though in Schutz's typification, the well-informed citizen does not aspire to expert knowledge whereas, of course, at least some sociologists do. The resemblance we find is in respect of what comprises commonly acceptable standards of knowledge and evidence. This is intensified when we consider how far sociological interests parallel the same order of socio-political concerns as engage what we normally think of as a well-informed citizen.

---

[37] 1962, to be exact. See *Explanation and Human Action.*

[38] For example, Jonathan H Turner lists the following 'areas of specialization':

History of Sociological Theory

Philosophy of Science

Contemporary Sociological Theory

Social Change

Theory Construction

Roles and Interaction

Comparative Institutions

Social Psychology

Social Stratification and Inequality

Historical Sociology

Race and Ethnic Relations

Human Social Ecology

Human and Societal Evolution

Sociology of Emotions

Bio-sociology

Modeling of Social Processes

[39] Defined by Schutz (op cit) as: 'the expert's knowledge is restricted to a limited field but therein it is clear and distinct. His opinions are based upon warranted assertions; his judgments are not mere guesswork or loose suppositions.'

[40] Concisely identified by Schutz (op cit) thus: "On the one hand, he neither is, nor aims at being, possessed of expert knowledge; on the other, he does not acquiesce in the fundamental vagueness of a mere recipe knowledge or in the irrationality of his unclarified passions and sentiments. To be well informed means to him to arrive at reasonably founded opinions in fields which as he knows are at least mediately of concern to him although not bearing upon his purpose at hand."

In the end, the conclusion we come to is this. Given the apparently irredeemable tension that exists between, on the one hand, the demands that the rigorous implementation of any serious model of scientific practice would make and, on the other, the desire on the part of sociologists to continue to talk about the topics they want to study and discuss in the ways they always have, it seems to us highly likely that any programme which is designed to impose unity through conformity to a strict model of science will be condemned to failure. This is a situation which will not be remedied by waiting for the sociological ugly duckling to emerge as a fully fledged swan but only when and if our professional colleagues decide the disciplinary leopard should change its cultural and organisational spots.

BIBLIOGRAPHY

Adler, J.	2009	'Resisting the force of argument.' *Journal of Philosophy*, pp 339-364
An, Li	2012	'Modeling decisions in coupled human and natural systems.' *Ecological Modelling*. No 229, pp25-36
Axelrod, R.	1997	*The Complexity of Co-operation*. Princeton. Princeton University Press.
Axelrod, R.	2006	'Agent-based modeling as a bridge between disciplines' in Tesfatsion, T. & Judd, K. (eds) *The Handbook of Computational Economics*, vol 2, Dordecht, Elsevier, pp 1566-1584
Axtell, R., Axelrod, R., Epstein, J., and Cohen, M.	1996	'Aligning simulation models'. *Computational and Amthematical Organization Theory*, vol 1 no 2, pp 123-141.
Baker, A.	2012	'Science driven mathematical explanation', *Mind*, vol 121, pp 243-267
Baldsarri, D.	2009	'Collective action' in Hedstrom, P. & Bearman, P (eds) pp 391-418
Bearman, P., Moody, J., & Stovel, K.	2004	Chains of Affection. *American Journal of Sociology*, vol 110, no 1: pp 44-91.
Bechtel, W. & Abrahamsen, A.	2005	'Explanation: a mechanist alternative'. *Studies in the History and Philosophy of Biological and Medical Sciences*, vol 36, pp 421-444
Berelson, B & Steiner, G.	1964	*Human Behavior: an inventory of scientific findings*. Oxford. Harcourt Brace.
Bernstein, B.	1971	*Class, Codes and Control, volume 1*. London . Routledge and Keegan Paul.
Biernacki, R.	2012	*Reinventing Evidence in Social Inquiry*. London. Palgrave Macmillan.
Blalock, H	1979	'Measurement and conceptualization problems'. *American Sociological Review*, vol 44 no 6 pp 881-894.
Blalock, H.	1979	'Dilemmas and strategies of theory construction' in Snizek, W., Fuhrman, E. and Miller M., (eds), *Contemporary Issues in Theory and Research*. London. Aldwyn Press
Bleda, M. & Shackley, S.	2012	'Simulation Modelling as a theory building tool.' *Journal of Artificial Societies and Social Simulation,* vol 15, no 2, 2
Boudon, R.	1998	'Social mechanisms without black boxes' in P. Hedstrom & R. Swedberg (eds), pp172-203
Breen, R.	2009	'Formal theory in the social sciences' in Hedstrom, P. &

		Wottrock, B. (eds) *Frontiers of Sociology*, Leiden, Brill, pp 209-230
Chase, I & Lindquist, W.	2009	'Dominance hierarchies' in Hedstrom, P and Bearman, P (eds), pp 566-591
Crowthers, C.	2013	'Thin Explanations'. *Philosophy of the Social Sciences,* vol 43, no 2, pp 257-267
Davidson, D.	2001	'Actions, reasons and causes', in Essays on Actions and Events, Oxford, Clarendon Press
Davis, J.	1994	'What's wrong with Sociology?' *Sociological Forum*, vol9 no 2, pp 179-197
de Grace P, and Stahl, L.	1990	*Wicked Problems and Righteous Solutions*, Englewood Cliffs, N.J. Prentice Hall.
Dean, J, Gumerman, G., Epstein, J., Axtell, R., Swedlund, A., Parker, M. % Carroll, S.	1999	'Understanding Anasazi culture change through agent-based modeling' in Kohler, T & Gumerman, G (eds) *Dynamics in Human and Primate Societies*. Oxford. OUP
Demeulenaere, P.	2011	'Causal regularities, action and explanation' in Demeulenaere, P (ed) *Analytical Sociology and Social Mechanisms*. Cambridge. CUP.
Douglas, M and Wildavsky	1983	*Risk and Culture*. Berkely. University of California Press.
Douglas, M.	1970	*Natural Symbols*. London. The Cressett Press.
Elster, J.	1989	*Nuts and Bolts*. Cambridge, CUP.
Elster, J.	2007	*Explaining Social Behaviour*. Cambridge. CUP.
Feld, S. & Grofman, B.	2009	'Homophily and the focused organization of ties' in Hedstrom, P and Bearman, P. (eds) pp 521-242
Franklin-Hall, L.	unpublished	*The Emperor's New Mechanisms* accessed from www.nyu.edu on 8/7/2013
Frigg, R & Reiss, J.	2009	'The philosophy of simulation', *Synthese*, vol 169, No. 3, pp 593-613
Galan, J. Izquierdo, L., Izquierdo, J., del Olmo ,R., Lopez-Paredes, A., and Edmonds, B.	2009	'Errors and artefacts in agent-based modelling.' *Journal of Artificial Societies and Social Simulation* vol. 12, no. 1, 1
Garfinkel, A.	1981	*Forms of Explanation*. New Haven. Yale University Press.
Garfinkel, H	1967	*Studies in Ethnomethodology*, Englewood Cliffs N.J. Prentice Hall.
Giddens, A.	1984	*The Constitution of Society*, Cambridge, Polity Press
Giere, R.	2010	'An agent-based conception of models and scientific representation.' *Synthese*, no 172, pp 269-281.
Gilbert, N.	2008	*Agent-based Models*. Thousand Oaks, CA. Sage
Gorski, P.	2009	'Social mechanisms and comparative historical sociology' in Hedstrom P. & Wittrock, B. (eds) *Frontiers of Sociology*. Boston. Brill
Gorski, P.	2013	'What is Critical realism? And Why Should You Care?"', *Contemporary Sociology*, vol. 42, no. 5, pp 658 - 670.
Gould, R.	1991	'Multiple Networks and Mobilization in the Paris Commune, 1871.' *American Sociological Review*, vol 56, no 6, pp 716-729.

Gould, R.	1993	'Trade cohesion, Class Unity and Urban Insurrection: artisanal activism in the Paris Commune.' *American Journal of Sociology*, vol 98, no 4, pp721-754
Hedstrom, P & Bearman, P.	2009	'What is Analytical Sociology all about?' in Hedstrom and Bearman (eds)
Hedstrom, P & Swedberg, R. (ed)	1998	*Social Mechanisms*. Cambridge, CUP
Hedstrom, P & Udehn, L.	2009	'Analytical Sociology and theories of the middle range' in Hedstrom & Hearman (eds)
Hedstrom, P & Ylikowski, P.	2010	'Causal mechanisms in the social sciences'. *Annual Review of Sociology*, vol 36, pp 49-67.
Hedstrom, P and Bearman, P (eds)	2009	*The Oxford Handbook of Analytical Sociology*, Oxford, OUP
Hedstrom, P.	2005	*Dissecting the Social*. Cambridge. CUP.
Hedstrom, P.	2008	'Studying mechanisms to strengthen causal inferences in quantitative research' in Box-Steffensmeier, J. Brady, H & Collier, D. *The Oxford Handbook of Political methodology*, Oxford, OUP, pp 319-338
Kaufmann, F.	1958	*The Methodology of the Social Sciences*. New Jersey, Humanities Press
Little, D.	1998	*Microfoundations, Methods and Causation*. New Brunswick. Transaction Publishers.
Louch, A.	1966	*Explanation and Human Action*. Berkeley & Los Angeles. University of California Press
Luhmann, N.	1996	*Social Systems*. California. Stanford University Press.
Lustick, I and Miodownik, D.	2009	'Abstractions, ensembles, and virtualizations.' *Comparative Politics* pp 223-244
Machamer, P., Darden, L., & Carver, C.	2000	'Thinking about mechanisms'. *Philosophy of Science*, vol 67, no 1, pp 1-25.
Marchione, E., Salgado, M and Gilbert, N.	2010	'"What did you say?"' Emergent communication in a multi-agent spatial configuration.' *Advances in Complex Systems*, vol 13, no 49, pp 469-482
Marchioni, C. & Ylikowski, P.	2013	'Generative explanation and individualism in agent-based simulation.' *Philosophy of the Social Sciences*, Vol XX no X pp1-18
Merton R	1948	'The self fulfilling prophecy'. *The Antioch Review*, vol 6, no 2, pp 193-210.
Merton, R.	1968	'On sociological theories of the middle range' in *Social Theory and Social Structure*, New York, The Free Press.
Midgley, D., Marks, R. And Kunchamwar,D.	2007	'Building and assurance of agent-based models. *Journal of Business Research*, vol 60, issue 8, pp 884-893
Moody, J.	2009	'Network dynamics' in Hedstrom, P. & Bearman, P. (eds) pp 447-476
Parsons, T.	1949	*The Structure of Action, Vol II*. New York, Free Press.
Polya, G.	1954	*Mathematics and Plausible Reasoning*, vols 1 & 2. Princeton. Princeton University Press.
Putnam, H.	1981	*Reason, Truth and History*. Cambridge. CUP
Reiss, J.	2009	'Causation in the social sciences'. *Philosophy of the Social Sciences*, vol 39, no 1, pp 20-40

Reynolds, C.	1987	'Flocks, herds and schools: A distributed behavioral model.', *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques* (Association for Computing Machinery): 25–34, doi:10.1145/37401.37406, ISBN 0-89791-227-6
Rorsch, E.	1977	'Human Categorisation' in Warren, N. (ed) *Advances in Cross Cultural Psychology vol. 1*. New York. Academic Press.
Salgado, M and Gilbert, N.	2012	'Emergence and communication in computational sociology'. *Journal for the Theory of Social Behaviour*, vol 43 issue 1, pp 87-110
Schelling, T.	1978	*Micromotives and Macrobehavior*. New York, W.W. Norton & Company.
Schutz, A.	1976	'The well informed citizen'. Collected Papers II . Dordecht. Springer
Schweder, R & Fiske, D.	1984	*Metatheory in Social Science*. Chicago. University of Chicago Press.
Sica, A.	2004	'Why "unobservables" cannot save General Theory'. *Social Forces*, vol 83, no 2, pp491-501
Sørensen, A.	1998	'Theoretical mechanisms and the empirical study of social processes,' in Hedstrom & Swedberg (eds) pp 238-66
Stinchcombe, A	1998	'Monopolistic competition as mechanism,' in Hestrom & Swedberg (eds), pp 267-305
Stinchcombe, A.	1986	Stratification and Organisation. Cambridge. CUP.
Turner, J & Kim, K-M	1999	'The disintegration of tribal solidarity among American Sociologists'. *The American Sociologist*, Summer, pp 5-20.
Vogt, P. And Coumans, H.	2003	'Investigating social interaction strategies for bootstraping lexicon development' , *Journal of Artificial Societies and Social Simulation* vol. 6, no. 1
Wittgenstein, L.	1972	*Philosophical Investigations*. Oxford. Basil Blackwell.
Woodward, J.	2003	*Making Things Happen*, New York, OUP
Wynne, B.	1994	'Public understanding of science.' In *Handbook of Science and Technology Studies*, (eds) S. Jasanoff et al. Thousand Oaks, CA.: Sage.