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.

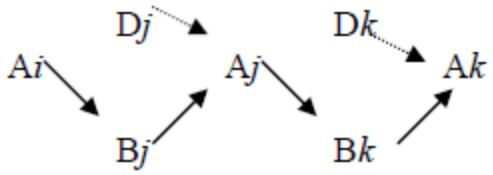
> *It is the year 1932. The Last National Bank is a flourishing institution. A large part of its resources is liquid without being watered. Cartwright Millingville has ample reason to be proud of the banking institution over which he presides. Until Black Wednesday. As he enters his bank, he notices 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.