**Introduction**

1. The principal motivator of Agent-Centred Sociology (ACS) is the simple observation that societies are societies *of agents*. All social actions are ultimately mediated by the actions of individual agents.
   1. This does not imply that all theories of social action (sociologies) need to explicitly concern themselves with the actions of agents any more than all physical theories need to concern themselves explicitly with the actions of atoms. There is a role for different levels of explanation. The possibility remains that there are valuable Higher-Level Sociological theories (HLSTs) whose ontologies do not include the agents themselves. Structural Sociology would be a possible example of such a theory.
   2. It *does* imply that no matter what ontology a sociology explicitly endorses, that theory must be *at least* consistent with a theory of the actions of these agents and their *local* causes and *local* effects. Moreover, a theory is not truly an *explanatory* theory until it provides a satisfactory causal relationship between antecedent conditions and consequent states. What makes a causal relationship ‘satisfactory’ in the required sense is a disputed question closely related to questions of the nature of understanding itself, but one way to address a failure of satisfactoriness in the causal relationship proposed between A and B is to argue that it is a plausible consequence of uncontroversial or more comprehensible or more ‘satisfactory’ causal relationships between the parts that constitute A and B. Where HLSTs, then, fail to be satisfactory, that failure may be addressed via some version of ACS – granted that the causal relationships referenced in ACS are held to be uncontroversial or more comprehensible or more satisfactory relationships than those in the HLST.

* 1. It suggests that valuable insights into the causes, processes, and effects of social action can be gained by investigating the causes, processes, and effects of action in individual agents. It may even be reasonable to aspire to the explanation of social actions and social characteristics entirely in terms of the actions and characteristics of individual agents, though we need not be committed to methodological reductionism.

1. Accepting the points above, and sometimes going further to claim, for example, that HLSTs cannot in principle constitute scientific knowledge, several kinds of ACS have been proposed. Rational Choice Theory (RCT), Social Exchange Theory (SXT), Symbolic Interactionism (SI), Ethnomethodology (EM), and Goffman’s work (EG) are the best known examples. In what follows, the infrastructure of an ACS of the kind adumbrated in work in RCT is described. This form is selected for the following reasons:
   1. RCT deals explicitly and directly with the actions of the agent, which is exactly what we are interested in, and the fundamental thesis of RCT is intuitively very plausible: in fact, with certain modifications it is almost a conceptual truth for intentional agents. The RCT thesis is that agents act to maximize their personal advantages, and the easy modification mentioned is to claim instead that the actions of an intentional agent are the practical consequences of rational deductions from the beliefs held by the agent directed at the satisfaction of its desires. We shall make this latter formulation the beginning point of the construction that follows.
   2. Amongst all ACS perspectives it can most plausibly be claimed to be a form of scientific investigation.
      1. It possesses characteristics which are advantageous to a system which aspires to the construction of explanations in the Deductive-Nomological (DN) style. For example:
         1. It deals directly with identifiable and testable causes and effects.
         2. It has a natural formalization.
         3. It lends itself to systematization.

Let it be noted that DN style explanations are not the only styles of explanation that have been called scientific: in particular, Inductive-Statistical explanations have been found acceptable, as have logical derivations. However, one may make two points concerning this: first, RCT is by no means excluded from developing explanations in those other styles; and second, a DN style of explanation in which the nomological elements are understood to describe causal relationships remains the gold standard for achieving satisfactoriness in understanding.

Let it also be noted that although the definition of DN arguments according to Hempel (1953, *Aspects of Scientific Explanation*) does not require causal laws, (1) that non-requirement was motivated by irrelevant-to-us philosophical concerns, (2) not requiring causal laws allows obvious non-explanations to pass as DN explanations, and (3) knowledge of causal relationships is essential to real understanding (as mentioned in 1.ii above.) In order to narrow the set of DN explanations more tightly around the set that increase our actual understanding, it would not be unreasonable to add to the list of criteria for DN explanations something along the lines that at least one of the minimal set of laws required for the explanation must be a causal law.

* + 1. The same characteristics that are advantageous in the construction of DN style explanations are also advantageous in the application of the Hypothetico-Deductive (HD) Method, which is the generally accepted method of science. The derivation of predictions/hypotheses to be tested in a formal system is much less fraught with difficulties of subjectivity, vagueness, etc. than such derivations in an informal system. And the testing of these hypotheses is much assisted by the following characteristic
* It lends itself to quantitative rather than qualitative research methods.
  1. It lends itself to varieties of approximation and simplification (particularly in the agent model) which make it relatively tractable as an analytic tool. And different approximations and simplifications can be made for different purposes.
  2. There are standard criticisms of RCT but, as we shall indicate in the course of the development, these can be met by relatively simple responses which do not adversely affect the advantages noted.
  3. It has had areas of success in application: for example, neo-classical economics and the various adaptations to the RCT which have been made to make the fundamental assumptions of that field more realistic have resulted in real advances. Other areas, such as political science, or criminology, also seem likely to be amenable to treatment by suitably modified versions of RCT.

1. ACS assumes the possibility of a model of an agent that is able to represent the relationships between local causes and effects and which is computationally tractable – or admits of reasonable approximations which are so. We will not attempt to construct a specific model of an agent, which is a task for actual scientists: however, in what follows the following assumptions are made:
   1. The model for the agent is a (simplified, obviously) model of processes that we believe are important in the decision procedure for any actual agent. In this respect we take an explicitly Realist position regarding the proposed models.
      1. Non-Realist approaches validate models that, in our view, are explanatorily impotent, whatever their predictive success.
      2. The Realist position puts a limit on the kinds of agent models that may be proposed that is responsive to work in other areas, such as social psychology, cognitive psychology, evolutionary psychology, neuroscience, etc.
   2. The agent model is restricted to just those aspects of the agent that are relevant to its role in the social world. This should not be interpreted as denying the importance in other respects of characteristics that are not deemed to be so relevant. Consciousness, for example, will be taken to be sociologically irrelevant (which does not mean that we must accept behaviourism in any form.)
2. The purpose of the following construction (plus commentary) is:
   1. To support the claim that the appropriate methodology for the social sciences and the nature of the explanations that are acceptable therein are of the same kind as in the ‘hard’ sciences. Ultimately, this is a claim that the understanding being sought in the social sciences is the same as the understanding being sought in the hard sciences, and that kind of understanding is one which is essentially of causal relationships.
      1. Referring to 1.ii, the considerations of 2.i-iii indicate that ACS is directed at the desired form of understanding of social facts, and it is therefore plausible that the causal relationships and ontology described in some ACS are of just the kind that would be required to address the failures of satisfactoriness identified in explanations presented in HLSTs. The causal relationships of ACS are generally thought to be uncontroversial or more comprehensible or more satisfactory, and the ontology is also generally thought to be uncontroversial. (At least, they are considered so compared to the elements of HLSTs – but there are plenty of philosophical questions surrounding all of these issues which are best ignored here.)
      2. Again, given the considerations of 1.ii, each version of ACS puts its own limits on the sorts of HLSTs that can be accepted. We will attempt to construct an ACS infrastructure which gives a guide to the kinds of HLST that may be accepted, by indicating the kinds of higher-level ontologies that can plausibly be constructed from the elements of this ACS. In particular, we will be concerned to indicate the extent to which sociological concepts which commonly occur in HLSTs, such as ‘interaction,’ ‘information,’ ‘identity,’ ‘dominance,’ etc., can be interpreted in terms amenable to treatment by an ACS; and to reduce to the degree possible the vagueness of such concepts and variables. The achievement of this latter goal will be much assisted by the strong formality of the ACS infrastructure constructed (c.f. 2.ii.a.2.)
      3. Let it be noted, however, that we do no not intend to indicate exactly how the micro-macro divide between ACS and HLSTs might be bridged.
         1. This would be an impossible task unless there was a particular HLST whose ontology and causal relations were taken to be the target, and we wish to make no commitments to any particular HLST.
         2. We are interested in providing evidence for the *plausibility* or the *possibility* of this sort of bridge, rather than in actually constructing the bridge.
   2. To clarify what the agent model needs to tell us about the agent in order for the ACS to be able to play the proposed limiting role on HLSTs.
   3. To address several criticisms of RCT in general (as well as SXT.) For example:
      1. In the basic form of RCT, there is an effective assumption of ‘perfect’ rationality in that it is assumed that an agent will always choose that which will maximize its preferences. It is said that this assumption of rationality is unrealistic and gives rise to paradoxes. It is further noted that in reality choices appear to be context-dependent in various ways (the framing of the choice, the emotional state of the agent, the presence of irrelevant choices, etc., all affect the outcome,) that much reasoning appears to be the application of heuristic processes (with the related observation that people just aren’t very good at reasoning anyway,) that the assumption of perfect information is far from accurate, and so on. All of this is valid criticism of the simplest RCT.

* We have already indicated (in 2.1) the modified form of theory that we will begin with, which meets some of these criticisms; nevertheless we still have a modified assumption of perfect rationality modulo the beliefs and desires of the agent. Without making any assumptions concerning the agent model, we will indicate at which points of the ACS infrastructure the assumption of perfect rationality may be corrected, and how this correction will propagate through the rest of the infrastructure.
  + 1. It is said that RCT assumes that the agent has a static character and that the social environment derives from the agent’s nature, whereas it is clearly the case that the agent’s character is shaped – at least in part – by the social environment. This latter relationship is in turn emphasized by SI, EM, and EG, which may be criticized in their turn for ignoring the fact that that the social environment is the result of the actions of the agents and these actions are derived – at least in part – from the agents’ characters.
* We will indicate various ways in which the social environment can affect the agent character, as well as vice versa.