Scientific Explanation and Human Emancipation

Roy Bhaskar

1. Introduction

What connections, if any, exist between explanations in the human sciences and the project of human emancipation? I want to address this issue in the light of the transcendental realist reconstruction of science (2) and the critical naturalism which that reconstruction enables (3).

My main target will be positivism, and the doctrine of the value-neutrality of social science. But I will also be attacking a rationalistic intellectualism, which sees social theory as (actually or potentially) immediately efficacious in practice.

In opposition to positivism, and its historicist/hermeneuticist displacements, I want to argue that the human sciences are intrinsically critical and self-critical; that accounts of social objects are not only value-impregnated, but value-impregnating; and that the possibility of a scientific critique of lay (and proto-scientific) ideas, grounded in explanatory practices based on respect for the authenticity and epistemic significance of those ideas, affords to the human sciences an essential emancipatory impulse, in virtue of which, subject to the operation of various ceteris paribus clauses, we pass securely from statements of fact to value.

However, in opposition to the idealist (theoreticist) notion of the unmediated efficacy of social science, I want to insist that it always occurs in the context of a situation co-determined by non-cognitive features too. Social theory appears, then, as conditioned critique: as subject, in its genesis and effect, to non-theoretical, as well as theoretical, determinations (whose critical understanding is itself part of the task of theory). This is of course an implication of historical materialism. To conceive critique as conditioned by factors outside itself is not to impugn its normative power, merely to be realistic about its practical impact.

On the view advocated here, knowledge, through necessary, is insufficient, for freedom. For to be free is (i) to know, (ii) to possess the opportunity and (iii) to be disposed to act in (or towards) one's real interests. Freedom can thus be no more the simple recognition of, than it is escape from necessity. Hegel (and Engels) and Sartre (and perhaps Marx, at least in his more chiliastic proclamations) are equally wrong - on the condition that circumstances or wants contain any non-cognitive components. It is salutary to remember that there is a logical gap between 'knowing' and 'doing', which can only be bridged by 'wanting in suitable circumstances'. It is the argument of this paper that the special qualitative kind of becoming free or liberation, which is emancipation, and which consists in the transformation, in 'self-emancipation' by the agent(s) concerned, from an unwanted to a wanted source of determination, is both causally presaged and logically entailed by explanatory theory, but that it can only be effected in practice.

2. Explanatory Schemata and Transcendental Realism

To explain something is to resolve some agent's perplexity about it: it is to render the unintelligible intelligible - by the elucidation, extension, modification or replacement of that agent's existing conceptual field (4). In particular, scientific explanations do not resolve problems by subsuming some particular problem under a more general one, but by locating such (normally already generalised) problems in the context of a new cognitive setting; it is (new) concepts, not (universal) quantifiers which accomplish explanatory problem-resolution in science. But the empirical adequacy of any such resolution must be tested by devising or finding conditions under which the referent of the (conceptual) object posited in the explanans operates free from extraneous influences. Now the enduring and transfactually active nature of such referents is a condition of the intelligibility of this experimental/exploratory activity; and so the philosophy of science must draw ontological distinctions between structures and events (the domains of the real and the actual) and open systems and closed, indexing the stratification and differentiation of reality (5).

Typically, then, to explain an event, regularity etc. is to bring it under a new

scheme of concepts, designating the structures, generative mechanisms or agents producing it. But, in line with their undifferentiated ontology, the dominant traditions in the philosophy of science have not clearly distinguished theoretical from practical (concrete, 'historical' or applied) explanations, neither of which are either deductive or inductive in form. Theoretical explanations are iteratively analogical - retroductive: i.e. antecedently available cognitive resources are used to construct plausible models of the mechanisms producing identified patterns of phenomena, which are then empirically checked out, and, if deemed adequate, in turn explained - in a continuingly unfolding dialectic of taxonomic and explanatory knowledge (6). Practical explanations involve the RRRE schema: i.e. resolution of complexes ('conjunctures' or 'compounds'), redescription of their components, retrodiction to possible antecedents of these components and elimination of alternative possible causes (7). Thus if theory assumes the form of the abduction of the abstract from the concrete, applied work characteristically depends upon the reverse movement, leading to the recovery from the abstract of the concrete, now reconstructed as the product of a multiplicity of abstractly apprehended determinants. Knowledge of structures and of their contingent modes of articulation in time thus appear as distinct moments of scientific activity. Between abstract sciences and the reconstructed concepts of concrete objects, lie the concrete sciences (like biography) which study the ensemble of significant truths about a given thing and the intermediate sciences (like ecology) which study the confluence of two or more orders of determination. Of course in as much as these types of explanation succeed in identifying real, but hitherto unrecognised, conditions and patterns of determination they immediately augment our knowledge, and hence (on the definition enlisted above) ceteris paribus our freedom.

On the metaphysics implied by the new analysis of science, ontology is vindicated as a study of the presuppositions of scientific practice, and the error of its reduction to epistemology is isolated. Moreover the world, as we actually know it (i.e. under the descriptions currently available to science), is now revealed as characterised by situations of dual and multiple control and by the phenomenon of emergence.

But transcendental realism does not license the simple-minded transapplication of results derived from reflection on the conditions of the natural sciences to the social sphere. Rather, it is only in virtue of an independent analysis, that we are in a position to see that there is a paramorphic relationship between the natural and the human sciences, such that there are knowable structures at work in the human domain partially analogous, but irreducible, to those identified in nature. Thus the material causality of social forms appears as a condition of intentional agency, and the efficient causality of beliefs as a condition of discursive thought. But a realist interpretation of non-physical (sui generis sociological, psychological) explanations of human phenomena is only justified if it can be shown that there are properties instantiated in the human world inexplicable in terms of different sets of conditions of purely natural laws. In concrete terms, the emergence of society is manifest in the causal irreducibility of social forms in the genesis of human action (or being), and the emergence of mind in the causal irreducibility of beliefs in the explanation of those changes in the states of the physical world which are the result of intentional agency. (Of course the relations are two-way. But the human effects of natural causes are normally mediated as cultural products, and the social effects of human actions in institutions.)

The resulting critical naturalism has nothing in common with either positivism or scientism, because clear differences transpire between positivism and science, on the one hand, and the human and natural sciences, on the other. Nor is it 'objectivist' in either method or result: for it is predicated on the analysis of (existing conceptualisations of) historical practices, and it situates these analyses within the framework of the same historical processes which social science describes and philosophy explicates. But positivism's anti-scientific hermeneutical foil is shown to be equally untenable - for the very features it picks upon (such as verstehen) themselves require for their intelligibility crucial aspects of the categorial framework of natural science (existential intransitivity, causality etc.). Nor do neo-Kantian syntheses of dual criteria or multiple interests fare any better. This is not only because the components of the attempted syntheses are faulty (e.g. in being based on a positivistic misconception of natural science), but because the very project of rendering ontological mediations as epistemological divisions is fundamentally mistaken. Thus conceptuality is a condition of generality in the historical domain; and there too an emancipatory conatus is initiated as an effect of explanatory power, in circumstances where it cannot be a universal or constitutive condition for it. (The critical cutting edge that Habermas' work retains despite this is achieved only by the effective noumenalisation of discourse as a counterfactual counterpoint to the realm of historical agency (8).)

3. Social Structure and Human Agency

On the transformational model of social activity (TMSA), entailed by the new critical naturalism, the ontological structure of human activity or praxis is conceived, after Aristotle, as consisting in the transformation by efficient (intentional) agency of pre-given material (natural and social) causes. A criterion for differentiating the social from the purely natural material causes is given by their property that, though necessarily pre-given to any particular agent, and a condition for every intentional act, they exist and persist only in virtue of human agency. On this model, then,

social structure and human agency are seen as existentially interdependent but essentially distinct (9). Society is both ever-present condition and continually reproduced outcome of human agency: this is the duality of structure (10). And human agency is both work (generically conceived), i.e. (normally conscious) production and reproduction of the conditions of production, including society: this is the duality of praxis. Thus agents reproduce, non-teleologically and recursively, in their substantive motivated productions, the unmotivated conditions necessary for - as means of - those productions; and society is both the medium and result of this activity. From this model flow a series of limits on naturalism, which may be summarised as the activity-, concept- and space-time-dependence of social forms, in virtue of which (as I have attempted to argue elsewhere (11)) a sui generis social science is possible. Of course the holistic, hermeneutical and historical character of social objects necessitate differences in the structure of social scientific explanations so that, paradigmatically, social complexes must be understood as partially conceptually articulated totalities in continual transformation. Similarly, the impossibility of artificially producing, and the unavailability of spontaneously occurring, closed systems requires reliance on purely explanatory (non-predictive) criteria of confirmation and falsification, and more generally theory-development and -assessment. However, in relation to the specificity of social objects, (non-scientistic) scientific knowledge of them is possible.

The TMSA allows us to pinpoint a double set of paired mistakes: the ontological errors of reification and voluntarism, and the epistemological ones of (social) determinism and (methodological) individualism. (Both may be combined to produce various pseudo-dialectical hybrids.) And it allows us to isolate the closely affiliated weaknesses of the substantive traditions of structuralism and functionalism, on the one hand, and action-oriented and interpretative sociologies, on the other. For its part, the TMSA respects a methodological distinction between the social sciences, which abstract from human agency, studying the structure of reproduced outcomes; and the social psychological sciences, which abstract from reproduced outcomes, studying the rules governing the mobilisation of resources by agents in their everyday interaction with one another and nature. If the object of the former is social structure, that of the latter is social interaction. They may be linked by the study of society as such, identified as the system of relations between the positions and practices agents reproduce and transform, the subject matter of the social science of sociology. The TMSA can allow that the form of psychology, the study of mental processes, may be species-general, but its content will always be historically specific.

The transformational model and the structure/praxis connection are represented as in Diagrams 1 and 2 below
On the TMSA unintended consequences, unacknowledged conditions and tacit skills

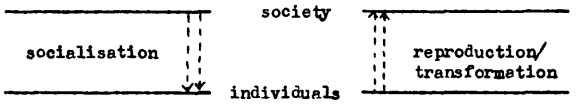

Diagram 1: The Transformational Model of Social Activity

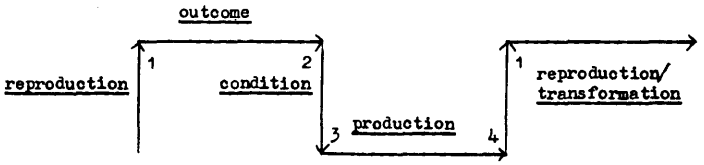

Diagram 2: Structure and Praxis

nb: 1, 1′ = unintended consequences; 2 = unacknowledged conditions; 3 = unacknowledged motivation; 4 = tacit skills

(cf. 1, 2 and 4 in Diagram 2) limit the actor's understanding of his social world, while unacknowledged (unconscious) motivation (cf.3) limits his understanding of himself. Corresponding to each of these limits, knowledge has a distinct emancipatory role - at 2 and 3 via the conditions and at 1 and 4 via the effects and form of praxis.

Now the continuity, depth and reflexivity of human agency suggest the model of it represented in Diagram 3, based on a model proposed by Anthony Giddens (12).

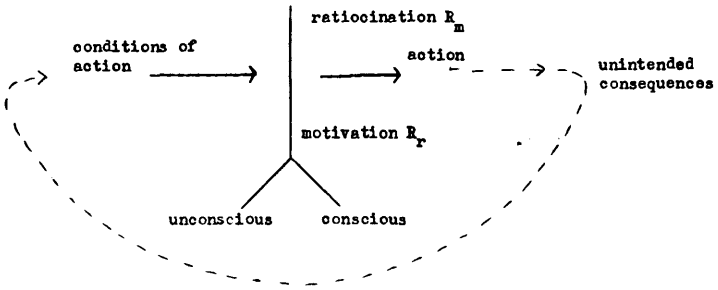

Diagram 3: The Stratification of Action

Discursivity presupposes a distinction between real and possible, including ratiocinated, reasons, grounded in the causal efficacy of the former. Ratiocination, R_m, is a property of the reflexive monitoring of conduct. Where $R_m \neq R_r$ there is the possibility of rationalisation. Real reasons are the wants that prompt motivation and ceteris paribus issue in action (13). As such they may be regarded as efficacious beliefs, which may be conscious or unconscious, trained on objects of desire. And as such they consist in a cognitive-conative vector or perhaps better ensemble (see Diagram 4).

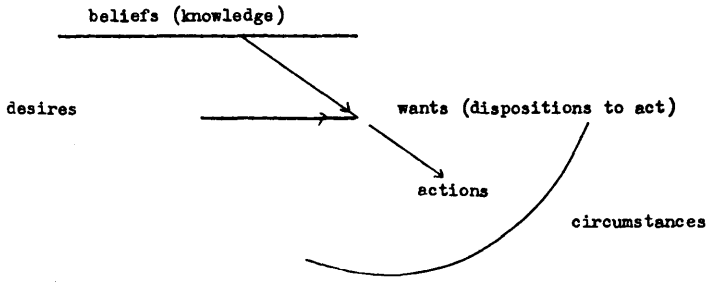

Diagram 4: Beliefs, Desires and Actions

The error of 'theoreticism' (see Section 1) is now clear: it involves the attempted elimination of the conative component, no doubt as rooted in our 'inner nature' as the cognitive component, from the generative matrix of action.

In social phenomenology unintended consequences may take the well-known forms of counterfinality and suboptimality (e.g. in a prisoner's dilemma) (14). The conditions figuring on the left-hand side of Diagram 3 include the rules and resources agents command in such games. Like all such conditions they may be unmotivated and unacknowledged. These are features that a general social phenomenology, whether rational or empirical, cannot itself, without vicious circularity, explain. For such rules and resources are at once historical deposits, and so always subject to a potentially unrecognised possibility of supercession. The games of the life-world (lebenswelt) are always initiated, conditioned and closed outside the life-world itself.

4. On the Critique of Interpretative Fundamentalism

The TMSA shows what may escape (and so be misconstrued by) consciousness in our conscious activity. But are there perhaps elements in our experience or aspects of our consciousness of which we must be certain, and which (perhaps in virtue of this) are not subject to the possibility of historical supercession?

The history of post-Cartesian philosophy is largely the history of the attempt to establish just such an Archimedean point for knowledge, free from the possibility of error and impervious to every form of doubt. Thus in a recent empiricist avatar, scientific knowledge was conceived as incorrigibly grounded in (or even exhausted by) sense-data or operations. Of course we now know that there are no foundations of knowledge, that there is no uniquely privileged level, moment or type of operation, that there are no brute data; that the facts already contain a certain 'sedimented' reading of the world (that natural facts are social institutions), and that the relationship between theories and facts is between the contents of two interdependent kinds of conceptual schemes, one of which is taken as referring to objects apprehended in experience. In short we now know that the facts are theory-dependent and changeable; and science itself appears, as one might anticipate on the TMSA, as a historical process of levels and connections, a weighted network, without foundations, developing in time. This view does not dispute the epistemic value of experience. However, it interprets this not as the absolute privilege of a content, but as dependent upon the ontological and social contexts within which the significant experience occurs (15).

Now in as much as there has been a 'coupure' in the recent philosophy of the human sciences, it lies in recognition of the significance of the condition that man is a self-interpreting and self-motivating animal, whose language and beliefs are in some manner necessary for and productive of his life; so that human reality faces the scientific neophyte as already pre-interpreted, as (as it were) linguistically and cognitively 'done', prior to any scientific investigation of it. These pre-interpretations are not externally related and contingently conjoined to what happens in the human sphere, but internally related to and constitutive of it (16). It was natural, then, in the wake of this understanding, to suppose that these interpretations (or beliefs) would constitute the base or foundations of social knowledge; to regard them as consisting, so to speak, in brute interpretations (or beliefs), whether such data-analogues were conceived positivistically as immediately available to the investigator or dialogically as dependent upon work within his own culture. Thus one had a transposition of the familiar thematics of classical philosophy in a hermeneutical key - more plausible than in the original, perhaps, because nature is not self-interpreting, but little different in logical form or epistemological effect. For both the reductionist thesis that social knowledge is exhausted by, and the milder position that it is rooted in (and so must be consistent with) self-interpretations lead inexorably to a displaced hermeneuticised scientism and a consequent 'disavowal of reflection' (17). In either variant the doctrine of the incorrigible, because ontologically constitutive, foundations of social knowledge secretes, like its positivist prototype, as an inevitable corollary: the doctrine of the neutrality of social science.

Of course Hegel demonstrated long ago (18) that the fundamentalist programme is both radically incomplete and viciously circular: in that it not only cannot establish its own legitimacy, but must (implicitly or explicitly) presuppose some unvalidated 'knowledge'. And it is clear that, in these respects, any Viconian facimus must share the same limitations as the Cartesian cogito. For just as Descartes must assume some content to initiate his axiomatics; so, for Vico, God or man must already possess some matter for their constructions, that is to make their worlds, and what any agent does not make (what it must take to make) it possesses no privileged understanding of (just as what an ego cannot demonstrate it must remain uncertain about). It should be noted that on the transformational model we do not make the conditions or consequences, skills or motives of our intentional making (cf. Diagram 2 above); so that our beliefs about, or interpretations of, our actions cannot be constitutive in the requisite sense.

In considering the social-incorrigibilist position in slightly more detail, it is convenient to distinguish two sub-arguments for it: one Viconian, the other hermeneutical in inspiration.

The more strictly Viconian argument contends that one and the same knowledge is used to generate as to explain behaviour; so, as it were, superimposing a transcend-

ental unity of agency on that of consciousness. But agency may consist in the exercise of tacit skills (19). Moreover, the consciousness involved, and knowledge exploited, in action may be practical and so cannot immediately ground, even if it is held to be the ultimate empirical touchstone of, a supposedly discursive, theoretical science. Thus we need not be able to say how we do what we know very well how to do (or vice versa), even when, as Chomsky has made abundantly clear, the first-order skills are themselves verbal, discursive ones. Secondly, while it is surely the case that communication (and interaction generally) would be impossible unless we were normally able to identify agents' immediate reasons for acting, it does not follow (a) that we must be always able to do so, or (more fundamentally) (b) that we must be able to identify the underlying reasons for (or causes of) those reasons. For example, we may know that a person is washing his hands or polishing an icon, but not why he is doing so. And so the possibility arises of the systematic misdescription of reasons in rationalisation or ideological-mystification, i.e. in the self-misunderstanding of agents or forms of life.

The hermeneutical argument for social foundations maintains that it is interpretations that uniquely and completely differentiate the social world from mere assemblages of physical happenings, so that it is only and sufficiently by reference to them that its sui generis character can be sustained. Elsewhere I have attempted to show that the social world is not exhausted by its conceptual aspects, and that such aspects are in any event not necessarily immediately available to consciousness (20). Thus although the immediate intentions of agents and meanings of acts cannot normally be misdescribed for mutual understanding or functioning language-games to be possible, both intentions and meanings may be opaque to agents (a) occasionally, at the level of everyday interaction and (b) systematically, at the level of the underlying explanations and descriptions of the reasons motivating their behaviour in such interaction. Particularly significant here is the possibility of a contingent generalisation of Gödel's theorem in the direction of what I shall call 'meta-critique'. This consists in a critique of a language on the grounds of its incapacity to adequately express ideas or institutions which are customarily described by means of them. Such a critique aims to pinpoint precisely what cannot be said in a particular language about what is said or done by means of it.

In general, then, the generative role of agents' skills and wants, and of agents' (and social) beliefs and meanings must be recognised without lapsing into an interpretative fundamentalism by conferring discursive and/or incorrigible status upon them. But how are beliefs and meanings in particular to be identified in the face of the corrigility of statements of them? Now agents' accounts are more than just evidence; they are an internally related aspect of what they are about. Thus any resolution of this problem must be two-way: the social investigator must avoid both the extremes of arrogant dismissal of and of fawning assent to first-person accounts (21). But agreement between agent and investigator hardly seems either a necessary or sufficient criterion for an adequate interpretation. Rather, it would seem that the adequacy of any interpretation, or more generally of any set of self-understanding, can only be shown in relation to the point of the interpretation (or understanding) in the always more or less contingently circumscribed context of the agents' self-formation, that is total developing life-activity (22).

If judgments about belief cannot be separated from judgments about activity, judgments of meaning, again presupposing a dialogical fusion of horizons (23), cannot be separated from judgments of explanatory adequacy (presupposing a degree of causal interaction). Thus the so-called 'problem of the indeterminacy of translation' is resolved in practice by selecting that translation which is explanatorily most adequate (whether or not it is the most 'charitable') in the context of what is already known about the organisation of the particular society in question (and societies in general). The most adequate explanation will save the maximum of significant phenomena in the subject matter at stake, showing in that subject matter precisely the degree and type of 'irrationality' that does so.

5. Facts and Values: Hume's Law and Helices

I now want to show that the human sciences are necessarily non-neutral; that they are intrinsically critical (both of beliefs, and of the objects of beliefs), self-critical and value-impregnating; and in particular that they both causally motivate and logically entail value-judgments ceteris paribus (CP). I will not be concerned to argue against the scientistic misconception that factual judgments are value-free, partly because this connection has been, if not always adequately theorised, widely recognised (inside as well as outside the analytic tradition (24)), but mainly because I want to address myself more to an aspiration than what is characteristically misconstrued as a 'difficulty': the hope that the human sciences might yet come to be in a position to cast some light on what we ought to do and say, feel and think.

In fact of course one is dealing with a fact-value helix here (see Diagram 5).

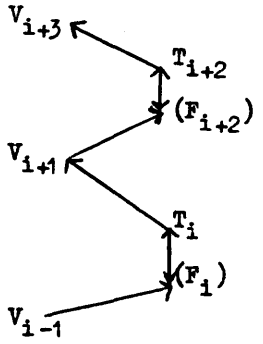

Diagram 5: Fact/Value Helix

And it is clear that the scientistic denial of the value-impregnation of factual discourse, involving the reification of propositional contents, shares with the positivist denial of its converse, as a common condition of their plausibility, a naive extensionalist theory of meaning (whether in physicalist, sensationalist or Platonist guise). Moreover it shares with the theoreticist (rationalist) conception of the unmediated efficacy of theoretical discourse a neglect of the conative and affective bases of action, involving a voluntarism of theoretical praxis. The converse 'practicalist' error - of anti-intellectualist irrationalism - ignores of course the cognitive bases of action. These four errors can be represented as in Table 1 below.

$F \not\rightarrow V$	positivism (and displacements)
$V \not\rightarrow F$	scientism
$T \not\rightarrow P$	irrationalism
$T \Rightarrow P$	theoreticism (idealism) $\longrightarrow P \not\rightarrow T$

Table 1

nb F stands for facts and theories
P stands for practice

Theoreticism, as defined above, leads naturally to the denial that practice (to the extent that it is not merely a redescription of 'theory') possesses any efficacy in the generation of theory.

Once the value-implications of theory, and the rational assessability of wants (in virtue of their grounding in beliefs), are accepted, then Diagram 4 can be modified as in Diagram 6.

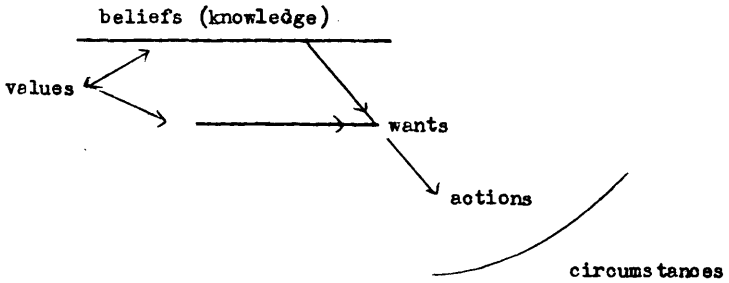

Diagram 6

Of course there is a feedback between values and actions, mediated by practices, including scientific (knowledge-producing) ones, so that they should be understood as connected by a loop as in Diagram 3.

There is an important asymmetry between the F → V and T → P relationships, on the one hand, and the V → F and P → T relationships, on the other. Factual and theoretical considerations not only predispose and motivate but, in favourable circumstances (and subject to the operation of CP clauses), logically entail value and practical judgments. On the other hand, value and practical considerations, while they may (and in general will) predispose and sometimes motivate, do not (non-trivially) entail factual and theoretical judgments (25). It is just this asymmetry which makes the helices in Diagram 5 (and in its theory/practice analogue) potentially relationalones: that is, progressive, i.e. developing, spirals, rather than viciously self-confirming, and so self-destroying, more or less rapidly vanishing, circles.

My core argument is very simple. In turns on the condition that the subject matter of the human sciences includes both social objects (including beliefs) and beliefs about those objects. Philosophers have characteristically overlooked, or concealed, the internal relations connecting these aspects: empiricists by objectivising beliefs, idealists by bracketing away objects. Now these relations, which may or may not be intra-discursive (depending upon whether the first-order object is itself a belief), are both causal and cognitive - in the ontological or intransitive dimension we are concerned with relations of generation; in the epistemological or transitive dimension of critique. But it is the causal relation of generation that grounds the epistemological programme of critique.

Now I am going to contend that if we possess (i) adequate grounds for supposing that a belief P (about some object O) is false and (ii) adequate grounds for supposing that S (co-) explains P, then we may, and must, pass immediately to (iii) a negative evaluation of S(CP) and (iv) a positive evaluation of action rationally directed at the removal of S(CP). To elaborate, in as much as we can explain, that is show the (perhaps contingent) necessity for some determinate false consciousness, or perhaps just some determinate consciousness under the determinable 'false', then the inferences to a negative evaluation of its source(s) and a positive evaluation of action oriented towards their dissolution are ceteris paribus mandatory.

It should be stressed straight away that such action can only be rationally justified CP to the extent that there are grounds for supposing the source to be dissoluble; and that the TMSA does not in itself license the supposition of a society without some false consciousness. The notion of false consciousness here involves simply in the first instance the notion of disjuncture, mismatch or lack of correspondence between belief and object. But, as I shall presently show, this general pattern of argument may be readily extended to accommodate both the cases of more interestingly specific forms of false consciousness and that of other types of inadequate consciousness (and, indeed, more generally, defective being).

In principle this pattern of inference applies equally to beliefs about natural, as well as social, objects, on the condition (and to the extent) that the relevant source of false consciousness S, is itself a social object. But in this case S cannot be the same as, or internally related to, O, and neither S nor P can be causal conditions for the genesis or persistence of O, as in the cases of psychological rationalisation and ideological mystification, where S, P and O are typically causally interrelated. Only

in the case of beliefs about social objects can the illusory (or more generally defective) character of consciousness be a condition of what it is about. However, given that beliefs about nature are social objects all the modalities of false consciousness may clearly apply to our beliefs about our beliefs about nature: that is, to our understanding of - as distinct from in - science.

I shall call (i) the critical and (ii) the explanatory condition. Of course even if the critical condition alone is satisfied, then we also pass immediately to a negative evaluation of P(CP), and of action based on or informed by P(CP). But I want to distinguish this kind of 'criticism' which, although it formally violates and so refutes 'Hume's Law' (26), remains silent on the causes of error, from an explanatory critique. Criticism, in Marx's words, 'knows how to judge and condemn the present, but not how to comprehend it' (27). The essence of Marx's objection to criticism may, I think, be stated thus: it employs value (and particularly, although contingently, moral) terms in the absence of any kind of causal grounding. At its best, i.e. if displayed in naturalistic (i.e. non-intuitionist or -emotivist) form, it can furnish grounds for belief and action which, if true, a fortiori increase our freedom. But criticism says nothing about, although it may of course (intentionally or unintentionally) causally affect, the (causal) conditions of actions, the springs (so to speak) of belief and behaviour, the sources of determination. And so criticism cannot contribute to the discursive elucidation of the topic of the transformation of the sources of an agent's determination from unwanted to wanted ones: i.e. of emancipation. Only a discourse in which the explanatory, as well as the critical, condition is satisfied can be intrinsically emancipatory.

As the concept of a 'critique' is better-known I shall not discuss it here. The structures of the various types of 'depth-explanation', which may be undertaken at several different levels (including the analytical, phenomenological and ideological), is considerably more complicated than that depicted in the bare form of an explanatory critique, but the transition from fact to value is effected in essentially the same way. The possibility of an explanatory critique constitutes the kernel of the emancipatory potential of the human sciences. But to illustrate the possibilities here fully, I want to develop the argument on a series of levels, which may be regarded as so many ratchets of reason.

6. Instrumental vs. Critical Rationality

At the first two levels, no attempt is made to question the logical heterogeneity (and impenetrability) of facts and values. Despite this, the human sciences may still have emancipatory implications (contingently, so to speak) in virtue of (i) their use as sheer technique and (ii) their effects in the context of the existence of relations of domination, exploitation and oppression.

Level I: Technical Rationality

Patently, the human sciences may be used, like any other sciences, to achieve (more or less consciously formulated, and justified) ends, which may of course be adjudged equally good or bad. In particular, explanatory theories may be used, in conjunction with statements of particular initial conditions, to generate technical imperatives akin to 'put anti-freeze in the radiator (if you want to avoid it bursting in winter) CP'. If such imperatives ever appear to depart from the ends-means schema, it is only because they already presuppose a context of human purposes in the domain of their intended applications.

Level II: Contextually-Situated Instrumental Rationality

The human sciences, even at the level of instrumental rationality, are not symmetrically beneficial to the parties involved in relations of domination etc. For, in the first place, explanatory knowledge increases the range of real (non-utopian) human possibilities, which may mean of course decreasing the range of assumed or fancied ones. But CP this will tilt the 'balance of - in a broad sense - political argument' against the status quo. This is quite consistent with the existence of only a simple external connection between knowledge and politics.

Secondly, even on an instrumental interpretation, explanatory knowledge appears as a necessary condition for rational self-emancipation (whether from the oppression of individuals, groups, classes, organisations, systems of relations, structures of interaction etc. or from the oppression of conscious or unconscious systems of ideas, in which the agent is entrapped). Hence the dominated, exploited, oppressed, repressed, etc. have an interest in knowledge (in the straightforward sense that it facilitates the achievement of their wants). And the dominating, in as much as their interests are antagonistic to those they dominate, possess an interest in the ignorance of the dominated (and perhaps even in their own ignorance of the nature, or even the fact, of their dominance). Thus the human sciences, and at a remove philosophy, cannot be regarded as equally 'a potential instrument of domination' as of 'the expansion of the rational autonomy of action' (28). The human sciences are not neutral in their consequences.

Level III: Intra-Discursive (Non-Explanatory) Critical Rationality

The point has been made, particularly effectively by Roy Edgley (29), that any science involves intra-discursive criticism, i.e. criticism of other actually or possibly believed (and therefore potentially efficacious) theories, hypotheses, etc. Acceptance of some theory T entails, ceteris paribus, a series of negative evaluations: on theories etc. incompatible with it, on

beliefs such theories underpin, on actions they sustain or inform. Granted that 'X is false' does not just mean 'Don't believe (act on) X' it certainly CP entails it. It is only if one denied any ontological connection between beliefs and action, or theory and practice, that one might have grounds for supposing that a change in theoretical does not entail a change in practical judgments (CP). But denying such a connection makes practical idscourse practically otiose. Again, this point is consistent with a contingent relationship between a science and its subject matter; and it applies, quite indifferently, at the level of intra-discursive critical rationality, to all sciences alike. All the sciences, then, irrespective of subject matter, are intrinsically critical, and so evaluative.

Level IV: Explanatory Critical Rationality

All the sciences make judgments of truth or falsity on beliefs about their object domain. But the human sciences, in virtue of the distinguishing feature of their object-domain that it includes beliefs about inter alia social objects, also make (or at least entail) judgments of truth or falsity on (aspects of) that domain. And such belief/object correspondence, or lack of it, appears immediately as a legitimate object of social scientific explanation. However, in as much as the natural sciences are also concerned in their own substantive critical discourse not just to isolate and criticise, but to comprehend and causally explain, illusory or inadequate beliefs about the natural world, then they too, assuming the second-order standpoint of the intermediate science (in the terminology of Section 2) of the natural-sociology (or -psychology) of belief - in which natural science is seen as a resultant of natural and cultural determinants (30) - may come to explain false consciousness of nature at least partially in terms of human causes (e.g. faulty instruments, inadequate funds, superstition, the power of the church, state or corporations). In virtue of their explanatory charter, and in as much as they are in a position to give well-grounded explanations of false consciousness, then, the human sciences must, and the natural sciences may (mediately, via the natural-sociology of belief), arrive at value judgments on the causes, as well as the contents, of consciousness.

To recapitulate the central argument, then, if we have a consistent set of theories T which (i) shows some belief P to be false, and (ii) explains why that, or perhaps some such false (illusory, inadequate, misleading), belief is believed; then the inferences to (iii) a negative evaluation of the object S (e.g. system of social relations) accounting for the falsity of the belief (i.e. mismatch in reality between the belief P and what it is about O) and (iv) a positive evaluation of action rationally directed at removing (disconnecting of transforming) that object, i.e. the source(s) of false consciousness, appear mandatory CP. This could be represented, informally, in the inference scheme below as:

$$\text{I.S.1 (i) T P. (ii) T exp I(P)} \rightarrow \text{(iii) - V(S} \rightarrow \text{I(P))} \rightarrow \text{(iv) V}\emptyset_{-S} \qquad (31)$$

and we certainly seem to have derived value conclusions (CP) from purely factual premisses.

Now for some possible objections.

1. It might be objected that 'P is false' is not value-neutral. But if it is not value-neutral, then the value-judgment 'P is false' can be derived from premisses concerning the lack of correspondence, or mismatch, of objects and beliefs (in the object domain). Moreover as, assuming that such judgments are intrinsic to any factual discourse, we are nevertheless able to infer from them, together with explanatory premisses, conclusions of a type which are not intrinsic to every factual discourse (viz. those specified in (iii) and (iv)), we do have a transition here that goes against the grain of Hume's Law, however precisely that is supposed to be here interpreted or applied. On the other hand, if 'P is false' is value-neutral, then the inferences to 'P ought not be believed (CP)' and 'Don't believe (act upon) P (CP)' certainly seem inescapable.
2. The suggestion that science itself presupposes, or embodies commitment to, certain values, such as objectivity, openness, integrity, honesty, veracity, consistency, coherence, comprehensibility, explanatory power, etc. should certainly be welcomed - suggesting, as it does, that the class of the 'value-neutral' is as empty as that of Austin's original 'constatives' (32). But it does nothing either to rescue Hume's Law, or to deny the validity of inference-types (iii) and (iv), which turn on the special feature of the sciences of beliefs that commitment to truth and explanatory power entail the search for theories which will possess value-implications that cannot be regarded as conditions of, or as already implicit as anticipations in the organisation of, scientific-activity-in-general.
3. It might be maintained that, although inference-type (iii) is valid, (iv) is faulty, so that no commitment to any sort of action is entailed by the critical explanatory theory. But this is not so. For one can reason straight away to action directed at removing the sources of false consciousness, providing of course one has good ground for supposing that it would do so, that no ill (or sufficiently overriding ill) effects would be forthcoming that that there is no better course of action which would achieve the same end. Of course the inference scheme does not itself, conceived as a philosophical reconstruction, determine what such practical ('critical-revolutionary') action is: that is the task of substantive theory. Of course 'remove (annul, defuse, disconnect, dissolve, transform) sources of false consciousness' does not specify what the sources are, any more than 'lying is wrong' says which statements are lies.

Behind this objection, however, lie two considerations of some moment. First, the kind of theory underpinning (iv) may be

different from that informing (iii). Diagnosis is not therapy. We may know that something is causing a problem without knowing how to get rid of or change it. Secondly, an explanatory critique of this type does not in general specify how we are to act after the source of mystification (false consciousness) is removed. It focuses on action which 'frees' us to act, by eliminating or disconnecting a source of mystification acting as an unwanted source of (co-) determination, replacing that source with another wanted (or perhaps just less unwanted) one, so achieving (absolute or relative) liberation from one stream of constraints or compulsions inherited from, as the causalities (and casualties) of the past. But it does not tell us what to do, if and when (and to the extent that) we are freed. Thus emancipated action may, and perhaps must, have a different logical form from emancipatory action.

The human sciences, then, must make judgments of truth and falsity, in virtue of their explanatory charter. And these, in the context of explanatory theories, entail value-judgments of type (iii) and (iv). Mutatis mutandis similar considerations apply to judgments of rationality, consistency, coherence, etc. Thus I.S.1 can be generalised in the cognitive direction represented in I.S.2 below, where C(P) stands for the contradictory character of some determinate set of beliefs.

I.S.2 T>P. T exp C(P) → -V(S C(P)) → $V\emptyset_{-S}$

But the human sciences are of course not only concerned to explain what might be called 'cognitive ills'. Their manifest includes the explanation of the 'practical ills' of ill-health, misery, repression, etc.; and in between such ills and the cognitive ones, what might be called the communicative ills of deception (including self-deception), distortion, etc.

This indicates two further lines of consideration. First I.S.1 can be straightforwardly generalised to deal with the explanation of such non-cognitive ills, with a corresponding deduction of value-judgments, as in I.S.3 below, where I-H stands for ill-health.

I.S.3 T exp I-H. -V(I-H) → -V(S → I-H) → $V\emptyset_{-S}$

However, as will be immediately obvious, this deduction, despite its evident social and epistemic power, is now no longer from purely factual premisses, or from what is immediately or self-evidently constitutive of purely factual discourse. And so it cannot be used to achieve a formal refutation of Hume's Law. It is precisely on this rock that most previous attempts at its refutation, including Searle's notorious attempted derivation of an 'ought' from the rather tenuous institution of 'promising' (33), have broken. But further reflection shows another possibility here: namely that there are non-cognitive conditions, such as a degree of good health and the absence of marked asymmetries in political, economic and the other modalities of power, for discourse (including factual discourse) -in-general to be possible. If this is correct then a formal derivation of an 'ought' can proceed as in I.S.4 below:

I.S.4 T>P. T exp (I-H → I(P)) → -V(S → I-H) → $V\emptyset_{-S}$

Is there a sense in which I.S.1 and 2 are epistemically prior to their non-cognitive generalisations? Yes, in as much as empirically-controlled retroduction to explanatory structures always occurs in the context of, and typically (in science) assumes the form of, criticism of beliefs (consciousness) - scientific, proto-scientific, lay and practical.

7. Depth Rationality

Level V: Depth-Explanatory Critical Rationality

The most thoroughly explored applications of I.S.1 and 2 involve the phenomena of psychological rationalisation and ideological mystification. These phenomena are characterised by two distinctive features. First, a doubling of necessity between misrepresentation (P) and source (S); so that the, or some such, misrepresentation is not only causally necessitated by, but causally necessary for, the persistence or modulation, reproduction or limited (non-essential) transformation of its source. Secondly, an internal relationship between source (S) and object (O); so that the misrepresented object

is either the same as, or at least causally dependent upon, the source of the misrepresentation.

Thus, in a simple depth-psychological model, an agent N may misdescribe his real (i.e. the causally efficacious) reason, s, for some action, ψ, by p. If p is itself a contingently necessary releasing condition for ψ and s itself generates, in context, p then we have:

(5) $s \rightarrow p.\ sp \rightarrow \psi.$

To explain this we now posit a structure S such that ψ is (perhaps contingently) necessary for its persistence or modulation, as in

(6) $S \rightarrow (s \rightarrow p.\ sp \rightarrow \psi) \rightarrow S'.$

Given $s \neq p$ the deductions proceed as in I.S.1.

This paradigm may be easily extended to include 'outer' as well as 'inner' causes, including the self-mystification of forms of social life, or systems of social relations, in ideologies. Thus the contradictions which mystify Colletti (34) turn simply on the necessary co-existence in social reality of an object and a (categorially) false presentation of it, where it is the inner (or essential) structure of the object which generates the categorially false presentation (or appearance).

Schema (7) is isomorphic with (5):

(7) $E \rightarrow A.\ EA \rightarrow P;$

and (8) is isomorphic with (6):

(8) $R \rightarrow (E \rightarrow A.\ EA \rightarrow P) \rightarrow R',$

where E = essence, A = appearance, P = practices, and R, R' the modulated reproduction of the capitalist mode of production.

Are there any general conditions on the internal structure (E) of a self-reproducing system (T) which generates and contains within itself (i.e. T) a functionally necessary misrepresentation (A) of itself? It seems plausible to suppose that E must possess at least sufficient internal differentiation to justify attributing to it a 'spaltung' or split; and that if T is to be capable of endogenous (essential) transformation, rather than merely modulated reproduction, the split must constitute, or be constituted by, antagonistic (opposed) tendencies. But apart from the Colletti-style contradiction built into the notion of the system's misrepresentation of itself, it seems a priori unlikely that what the human sciences may empirically discover about the various structural sources of false consciousness will justify the application of a single, unified category of 'contradiction' to those structures. Instead one might conjecture a galaxy of concepts of contradiction, clustered around the core notion of the axiological indeterminacy generated by the logical archetype (together with the evaluative connotations this secretes). The specific concepts of contradiction would then achieve their individuation in the constraints they impose upon such indeterminacy and in their thematisation of its form.

Perhaps the most famous depth-explanation, Marx's Capital, has the structure of a triple critique: of theories, of the practical consciousness such theories reflect or rationalise, and of the conditions explaining such consciousness. But in Marx, and the Marxisn tradition generally, the criticised (discursive and practical) consciousness is regarded not just as false but as 'ideological' - where 'ideology' is counterposed to 'science'. In addition to the critical and explanatory conditions, one thus finds a further set of categorial conditions. Here beliefs are typically criticised for their unscientificity simpliciter, or for their inadequacy in sustaining the (irreducible) specificity of the subject matter of their domains. Thus in reification, fetishism, hypostatisation, voluntaristic conventionalism, organicism, etc. social life is presented, in one way or another, in an a-social mode - a condition rooted, for Marx, in the alienation and atomisation characteristic of capitalism as a specific form of class society. For example, on Marx's analysis, the wage-form collapses a power (labour-power) to its exercise (labour), the domain of the real to the actual, while the value-form fetishistically represents social relations in the guise of natural qualities. The critique of these gross categorial errors could be represented as:

I.S.9 $T>P.\ T \exp I(P).\ T \exp -S_c\ (P) \rightarrow -V(S \rightarrow -S_c.\ I(P)) \rightarrow V\emptyset_{-s};$ and

I.S.10 $T>P.\ T \exp I(P).\ T \exp -S_o\ (P) \rightarrow -V(S \rightarrow -S_o.\ I(P)) \rightarrow V\emptyset_{-s},$

where $-S_c$ and $-S_o$ stand for the unscientific and desocialising character of the forms in question.

What are we to make of Engels' celebrated rebuke to Lefargue: 'Marx rejected the "political, social and economic ideal" you attributed to him. A man of science has no ideals, he elaborates scientific results, and if he is also politically committed, he struggles for them to be put into practice. But if he has ideals, he cannot be a man of science, since he would then be biased from the start' (11 August 1884)? While interests both predispose and motivate analyses (and their acceptance/rejection) in the human sciences, so that Engels' scientistic repudiation of the V → F connection is disingenuous; it remains the case that no value judgments other than those already bound up in the assessment of the cognitive power of Marx's theory are necessary for the derivation of a negative evaluation of the capitalist mode of production (CP) and a positive evaluation of action rationally oriented towards its transformation (CP) - so that the political commitment that Engels attributed to Marx as, so to speak, a contingent extra, can (on the assumption that Marx's depth-explanation is correct) be logically grounded in his scientific practice alone. Of course the theories now required to confirm, extend, develop or refute Marx's own analyses can only be consequent upon engagement in invest-

igations of comparable scope and penetration.

Level VI: Depth Rationality

Given that clear paradigms exist in the human sciences of I.S.1-4, most notably in the traditions inaugurated by Marx and Freud but also in some of the work of the theorists of the life-world of social interaction, is there a sense in which the application of these inference schemes, and hence of the type of explanatory critique they presuppose, is transcendentally necessary?

Now assume two interlocutors X and Y. Suppose X believes himself to possess a rational argumentative procedure R_A, a reasoned argument A_r and a conclusion Q; but that Y does not or cannot (perhaps 'in spite of himself') accept or act upon R_A, A_r or Q. (The reverse conditions may apply symmetrically to X, but we can ignore this complication here.) What is to be done when rational argument fails? Clearly there are three general kinds of possibility here:

(i) Y continues to mistakenly believe (and act upon)-Q;

(ii) some non-discursive process (e.g. force, medication) induces in Y a belief in Q; or

(iii) X and Y jointly initiate an inquiry into the conditions blocking or compelling Y's beliefs.

Adoption of solution (i), i.e. stoic acceptance of irrationality, error, etc. is a counsel of despair. Moreover it cannot be generalised to the first person case of doubt (or more generally, choice) without vicious axiological regress. Solution (ii) can be ruled out on the grounds that drugs, force, etc. can only simulate the acceptance of A_r or R_A. Further it is not emancipatory, in that it does not replace an unwanted with a wanted source of determination, but merely counteracts the effects of one unwanted source of determination with another. This has the corollary that in as much as the original source of determination is not defused, it may continue to exercise a latent power.

The alternative (iii) of a depth-investigation (D-I) is possible where reason fails but has not yet exhausted its resources; and it is practicable where Y's beliefs are generated or underpinned by unreflected (unacknowledged) processes, and Y seeks to understand, in order to undermine or abrogate, these processes. A depth-investigation may be defined generally as any cooperative inquiry, which includes the agent, into the structure of some presumed set of mechanisms, constituting for that agent an unwanted source of determination (which, whether cognitive or not, will always possess some cognitive manifestation), with a view to initiating, preserving or restoring that agent's ability to act and think rationally.

Four points must be immediately made about this definition. First, what is rational cannot be stipulated a priori, but must itself be discovered, in relation to antecedent notions of rationality (its nominal essences, so to speak), in the context of the explanatory critique such a depth-investigation presupposes. Secondly, although the concept of a depth-investigation has been introduced as an ideographic practically-oriented application of some or other determinate explanatory critique, the theory at the heart of the critique itself depends crucially for its own development and empirical confirmation on such investigations (whether on living or reconstructed, e.g. historical, materials). It follows from this that the links between theory and practice, and between pure and applied research, though not abrogating their distinctions, are bound to be tighter than in the natural sciences. Thirdly, corresponding to the different types of inference scheme outlined above, there will be different forms of depth-investigation. These must not, however, by hypostatised. For of course the explanation of cognitive ills will in general involve reference to practical and communicative ills, and vice versa. Finally the desire for emancipation which motivates the depth-investigation can neither be posited a priori (for although it is a necessary truth that people act on their wants, it is not a necessary truth that they act on their interests), nor predicted in historicist fashion on the basis of some particular theory of individual development of history. But as a socially-produced social object, the desire for emancipation will of course be a crucial topic for meta-investigations. And such investigations will need to be continually reflexively incorporated into the substantive theory of the practice etc. from or for which emancipation is sought.

The structure of a simplified D-I may be elucidated as follows:

(1) Y is not capable of $\emptyset$; scientific realism suggests there is a mechanism M preventing this.

(2) General theory T investigates the

structure of blocking/compelling mechanisms, under the control of empirical data and researches.

(3) The application of T to Y investigates the agent Y, as well as X. For it is Y's interpretations, actions and determinations that are at issue. Subjectivity in the human sciences is not an obstacle; it is (an essential part of) the datum. But ontological authorship does not automatically carry over into epistemological authority. Now the Y-dependence of the D-I means that Y must have a motive or interest in disengaging M, or in a range of aspects that M prevents. And that co-investigator X must not have an interest in the distortion of M-descriptions. Concretely, this raises the questions of the costs of emancipation for Y and of the conditions under which emancipation may be a second-best solution for Y; and for X it presupposes both the willingness to learn (in the general spirit of Marx's 'Third Thesis on Feuerbach') and the continuing development of X's own self-understanding. At a deeper level, the success of the detailed investigation of the modus operandi of M in or for X must depend upon an internal differentiation within the experience of X, so that the empiricist/utilitarian notion of emancipation as a process of the alteration of the circumstances of atomistic individuals must be rejected. Moreover it should be reiterated that cognitive emancipation will in general depend upon non-cognitive (and extra-discursive) conditions; and that cognitive emancipation is necessary, but insufficient, for full emancipation (as shown by the example of the slave who knows very well he is a slave but still remains a slave, i.e. unfree).

In fact dissonance, not liberation, may be the immediate result of enlightenment. And such dissonance may lead either to 'revolutionary-critical' activity or to despair. Moreover constraints upon cognitive emancipation itself are imposed by the pre-formation of thought-contents (in psycho-analysis), the projects of others (in social phenomenology) and the non-discursive aspects of social reality (in historical materialism). Hence emancipation cannot be conceived either as an internal relationship within thought (the idealist error) or as an external relationship of 'educators', 'therapists' or 'intellectuals' to the 'educated', 'sick' or 'oppressed' (the empiricist error).

Now I want to propose that the possibility of a depth-investigation is a transcendental condition for any science of man and hence (at a remove) for any science at all; and that in particular to inquire into the nature of the real grounds for beliefs is the same thing as to inquire into the possibility of rationalisation, self-deception, deception of others, counterfinality and systemic mystification; and that to inquire into the conditions of possibility of these cognitive-communicative malaises immediately raises the question of the conditions of the possibility of practical ones - from ill-health to brutal oppression. The issue of the causes of belief and action, presupposing a distinction between real and possible (including assumed or fancied) grounds, can only be taken up the depth human sciences. But a moment's reflection shows that this distinction, and hence the possibility of a depth-investigation at the analytic, phenomenological and historical levels, is a condition of every rational praxis or authentic act of self-understanding at all. It is necessitated by the existential intransitivity and enabled by the causal interdependency of the phenomena of sociality. Thus in the human sciences the problem of error (oppression, etc.) must make way for the problem of the causes of error (oppression, etc.), as part of the programme, paramorphic (but non-identical) to that of Kepler, Galileo and Newton, of the investigation of the underlying structures producing the manifest phenomena of social life.

The object of the depth-investigation is emancipation. Emancipation may be conceived either as the process of the changing of one mode of determination D, into another D_2, or as the act of switching from D_1 to D_2, both D_1 and D_2 perduring but D_1 in an inactivated condition. Now if the emancipation is to be of the human species, then the powers of emancipated man must already exist (although perhaps only as powers to acquire or develop powers) in an unactualised state. The key questions for substantive theory then become: what are the conditions for the actualisation of the powers?: are they stimulating (of the socialist tradition); or releasing (of the anarchic/liberal traditions)?; do they lie in social organisation or individual attitudes etc? (35)

8. Conclusion

Can anything be said about the conditions of the possibility of emancipatory practices in general? I think that, for emancipation to be possible, four general types of condition must be satisfied.

First, reasons must be causes, or discourse is ontologically redundant (and scientifically inexplicable). But the potentially emancipatory discourse, given the TMSA and the general conception of an open world, can only co-determine action in an already pre-structured, practical and collective context.

Second, values must be immanent (as latent or partially manifested tendencies) in the practices in which we engage, or normative discourse is utopian or idle. I think that Marx, in conceiving socialism as anticipated in the revolutionary practice of the proletariat, grasped this. And it is on this feature that Habermas' deduction of speech-constitutive universals also turns (36). But if there is a sense in which the ideal community, founded on principles of truth, freedom and justice, is already present as an anticipation in every speech inter-action, might one not be tempted to argue that equality, liberty and fraternity are present in every transaction or material exchange; or that respect and mutual recognition are contained in the most casual reciprocated glance? (37). It is an error to suppose that ethics must have a linguistic foundation; just as it is an error to suppose that it can be autonomous from science or history.

Third, critique must be internal to (and

conditioned by) its objects; or it will lack both epistemic grounding and causal power. But it follows from this that it is part of the very process it describes, and so subject to the same possibilities, of unreflected determination and historical supercession, it situates. Hence continuing self-reflexive auto-critique is the sine qua non of any critical explanatory theory.

Finally, for emancipation to be possible, knowable emergent laws must operate (38). Such laws, which will of course be consistent with physical laws, will be set in the context of explanatory theories elucidating the structures of cognitive and non-cognitive oppression and the possibility of their transformation by women and men. Emancipation depends upon the untruth of reductionist materialism and spiritualistic idealism alike. On reductionism - if the physical process level is L_p, and the level at which emancipation is sought is L_e, then either L_p completely determines L_e and no qualitative change is possible; or qualitative change is possible, and the laws of L_p are violated. On idealism - either emancipation is entirely intrinsic to thought, in which case it is unconditioned and irrationality is inexplicable; or if it is conditioned, it cannot be intrinsic to thought. Emancipation depends upon explanation depends upon emergence. Given the phenomenon of emergence, an emancipatory politics or therapy depends upon a realist science. But, if and only if emergence is real, the development of both are up to us.

The possibility of emancipation is not of course the reason why an emergent powers theory, if it is, is true. It is rather that if human beings, and social forms in general, are emergent from but conditioned by nature, then there is at least the possibility that the human sciences, provided they 'do not anticipate the new world dogmatically, but rather seek to find the new world through criticism of the old' (39), could still be of some benefit to the greater majority of mankind.

Footnotes

1 I would like to take this opportunity to acknowledge some debts. First, I have benefitted greatly from discussions with Roy Edgley on this and related topics. Secondly, I owe much to the stimulus of the pioneering work of Jürgen Habermas in this field, even where (as will be obvious) I come to rather different conclusions. Thirdly, I am indebted to the writings of Alasdair MacIntyre, who did perhaps more than anyone else in the 'analytical' tradition to open up the possibility of a historical treatment of moral and practical philosophies. Finally, this present paper developed out of another 'Emergence, Explanation and Emancipation' presented at a conference organised by Paul Secord under the auspices of the University of Houston (and forthcoming in Conceptual Issues in the Human Sciences, ed. P. Secord, Oxford 1981). I am extremely grateful to the participants at that conference, and at the seminars where I have read drafts of this paper, for their criticisms, questions and comments; and in particular to William Outhwaite for sharpening my thinking on the nature of an explanatory critique of consciousness in natural science.

2 See my A Realist Theory of Science 1st ed. Leeds 1975, 2nd ed. Harvester Press, Brighton and Humanities Press, New Jersey 1978, and my 'Realism in the Natural Sciences', Logic, Methodology and Philosophy of Science VI, ed. L.J. Cohen et al, North Holland, Amsterdam 1980.

3 See my The Possibility of Naturalism, Harvester Press, Brighton and Humanities Press, New Jersey 1979.

4 Cf. e.g. M. Scriven, 'Truisms as the Grounds for Historical Explanation', Theories of History, ed. P. Gardiner, Free Press, New York 1959; and P. Achinstein, 'Explanation', American Philosophical Quarterly Studies in the Philosophy of Science, ed. N. Rescher, Blackwell, Oxford 1969.

5 See A Realist Theory of Science.

6 Cf. esp. N.R. Hanson, Patterns of Discovery, Cambridge University Press, Cambridge 1958; R. Harré, The Principles of Scientific Thinking, Macmillan, London 1970; and M.B. Hesse, The Structure of Scientific Inference, Macmillan, London 1974.

7 See The Possibility of Naturalism, p165.

8 See e.g. J. Habermas, Theory and Practice, Heinemann, London 1974, pp16ff.

9 Cf. A. Collier, 'Materialism and Explanation', Issues in Marxist Philosophy Vol.II, ed. J. Mepham and D.H. Ruben, Harvester Press, Brighton 1979, p37, and the unpublished essay by M. Westlake referred to there.

10 Cf. A. Giddens, New Rules of Sociological Method, Hutchinson, London 1976, p121 & passim.

11 See The Possibility of Naturalism.

12 A. Giddens, Central Problems in Social Theory, Macmillan, London 1979, p56.

13 See The Possibility of Naturalism, pp121-23.

14 See e.g. J. Elster, Logic and Society, Wiley, Chichester 1978, Ch.5, and E. Ullman-Margalit, The Emergence of Norms, Oxford University Press, Oxford 1977.

15 See A Realist Theory of Science, Ch.1.

16 Cf. e.g. C. Taylor, 'Interpretation and the Sciences of Man', Review of Metaphysics 25(3), 1971 (reprinted in Critical Sociology, ed. P. Connerton, Harmondsworth 1976).

17 J. Habermas, Knowledge and Human Interests, Heinemann, London 1972, p.vii.

18 G.W.F. Hegel, The Phenomenology of Mind, Allen & Unwin, London 1949.

19 Cf. M. Polanyi, Personal Knowledge, Routledge and Kegan Paul, London 1958, Ch.4.

20 See The Possibility of Naturalism, esp. Ch.4, Sections 3 and 4.

21 Cf. R. Bernstein, The Reconstruction of Social and Political Theory, p203.

22 Cf. Knowledge and Human Interests, passim.

23. Cf. H.-G. Gadamer, Truth and Method, Sheed and Ward, London 1975.

24 Cf. C. Taylor, 'Neutrality in Political Science', Philosophy, Politics and Society 3rd Series, ed.s P. Laslett and W. Runciman, Blackwell, Oxford 1969 (reprinted in The Philosophy of Social Explanation, ed. A. Ryan, Oxford University Press, Oxford 1973).

25 If 'ought' implies 'can', the non-trivial implication of a power is a presupposition, not an entailment, of the ought-statement; which depends upon a theory (i.e. 'factual knowledge') of the agent and his circumstances.

26 See R. Hare, Freedom and Reason, Oxford University Press, Oxford 1963, p108. The title has been (in my opinion, unconvincingly) disputed - see e.g. A. MacIntyre, 'Hume on "is" and "ought"', Philosophical Review 1959, reprinted in A. MacIntyre, Against the Self-Images of the Age, Duckworth, London 1971. (The vexed passage in Hume is A Treatise of Human Nature, ed. L.A. Selby-Bigge, Oxford University Press, Oxford 1965, III, i.1, pp468-70.)

27 K. Marx, Capital Vol.I, Lawrence and Wishart, London 1965, p505

28 A. Giddens, New Rules of Sociological Method, p159.

29 See esp. R. Edgley, 'Reason as Dialectic' Radical Philosophy 15, 1976, and 'Marx's Revolutionary Science', Issues in Marxist Philosophy Vol.III, eds. J. Mepham and D.H. Ruben, Harvester Press, Brighton 1979.

30 Cf. e.g. D. Bloor, Knowledge and Social Imagery, Routledge and Kegan Paul, London 1962, Ch.2.

31 An explanatory critique in the natural sciences could be represented as follows:

I.S.1' (i) T>P. (ii) T exp $I(P_N)$ → (iii) $-V(S_S \rightarrow I(P_N))$ → (iv) $V\emptyset_{-S_S}$

32 See J. Austin, 'Performative-Constative', La Philosophie Analytique 1962, trans. by G.J. Warnock in Philosophy and Ordinary Language, ed. C. Caton, University of Illinois Press, Urbana 1963 (see esp. p30).

33 J. Searle, 'How to derive "ought" from "is"', Philosophical Review 1964, and Speech Acts, Cambridge University Press, Cambridge 1969, Ch.8.

34 See L. Colletti, 'Marxism and the Dialectic', New Left Review 93 (1975).

35 Is the present direction of argument necessarily incompatible with a substantive utilitarian ethics? Yes and no. The utilitarian tradition has generally been willing to concede that a world with more possibilities is CP better than one with less, so that supposing that a happy or healthy man could make himself miserable or ill (not that he would - in virtue of his state - of course normally want to) but not vice versa, Bentham and Mill would be bound, on this kind of ground alone, to approve an emancipated as better than a non-emancipated state. But could they approve the kind of measures substantive depth theories indicate as necessary for the transition to such states? I doubt it. In Kantian terms, one could say that although they might will the end, it is highly unlikly that they could will the means in all but the most improbable circumstances.

36 See J. Habermas, 'Towards a Theory of Communicative Competence', Inquiry 13 (1970).

37 Cf. R. Harré, Social Being, Blackwell, Oxford 1979.

38 See my 'Emergence, Explanation and Emancipation', Conceptual Issues in the Human Sciences, ed. P. Secord, Oxford 1981.

39 See Writings of the Young Marx on Philosophy and Society, ed. Easton and Guddat, Anchor Books, New York 1967, p212.