

PSYCHOANALYSIS AND HUMAN EMANCIPATION*

Russell Keat[†]

In the Introduction I noted how Habermas presented psychoanalysis as ‘the only tangible example of a science incorporating methodical self-reflection’,¹ and thus as a model for the general epistemological features of a critical social theory guided by the emancipatory interest in human autonomy. I will now examine his account in more detail, especially its claims about the relations between autonomy, causality and the unconscious. By doing this I hope to show how certain features of this view of critical social theory can be preserved in a reconstructed form which is basically deterministic, and which builds upon some of my suggestions in the previous chapter about how to characterize the human species by comparison with other ‘objects’. Discussion of Habermas’s claim about the distinctive criteria of validity for critical theory, which is also developed through his account of psychoanalysis, will be left till the next chapter.

In the first two sections I argue that there are a number of defects in Habermas’s account of Freud, considered solely as an exegesis. These mainly concern his interpretation of the concepts of the id and the unconscious, and of the goal of psychoanalytic therapy. By itself this criticism is not, of course, particularly significant. The acceptability of Habermas’s view of critical social theory does not depend upon the exegetical accuracy of this account: for he might be wrong in thinking that psychoanalytic theory provides an epistemological model for critical social theory, without this undermining his conception of the latter. Further, we must remember that Habermas asserts that Freud misunderstood, scientifically, the epistemological status of his own theory, and that there are features of the theory which reflect this, and which must therefore be removed if psychoanalysis is to be a ‘pure’ example of a self-reflective science.² However, I shall argue that the exegetical deficiencies in Habermas’s account indicate serious problems for the view of critical social theory which he intends psychoanalysis to illustrate. And, in some respects at least, I think Freud’s ‘self-misunderstanding’ is preferable to Habermas’s. In particular, I try to defend in the last three sections a deterministic and

* [2013] Published as Chapter 4 of *The Politics of Social Theory: Habermas, Freud and the Critique of Positivism*, Basil Blackwell/University of Chicago Press 1981, pp. 94-132; citations should be to this. No changes have been made to the original text, apart from typographical corrections and the addition of bibliographical information. References in the text to other chapters in the book have been retained: chapter 2 (‘Value-Freedom and Socialist Theory’), chapter 3 (‘Knowledge, Objects and Interests’), chapter 5 (‘Theory and Practice in Psychotherapy’), and the Introduction, are also available at www.russellkeat.net; so, too, is an earlier version of chapter 1 (‘The Critique of Positivism’).

[†] [2013] School of Social and Political Science, University of Edinburgh: russell.keat@ed.ac.uk. Previously, Department of Philosophy, Lancaster University.

reductionist position which is generally consistent with Freud's, but which also enables something similar to Habermas's conception of psychoanalysis as an emancipatory process to be preserved.

1. The id: alienated ego or home of the instincts?

The structural model of id, ego and superego, says Habermas, is intrinsically related to the self-reflective character of the therapeutic process in psychoanalysis. In that process the nature and meaning of the distorted communications typical of neurosis are revealed by undoing the work of the unconscious which, through a complex series of defences, has disguised the wishes or desires earlier repressed from consciousness. Thus intelligibility is restored to the 'distorted texts' of neurotic activity, and the patient recovers what had been banished from the realm of public communication. I emphasize 'recovers', since this is crucial to Habermas's account: what was lost is found again. The Hegelian theme of reconciliation, of the overcoming of alienation, seems evident here. Thus Habermas says:

the insight to which analysis is to lead is indeed only this: that the ego of the patient recognize itself in its other, represented by its illness, as in its own alienated self and identify with it. As in Hegel's dialectic of the moral life, the criminal recognizes in his victim his own annihilated essence; in this self-reflection the abstractly divorced parties recognize the destroyed moral totality as their common basis and thereby return to it.³

And, rather less evocatively:

Analysis has immediate therapeutic results because the critical overcoming of blocks to consciousness and the penetration of false objectivations initiates the appropriation of a lost portion of life-history; it reverses the process of splitting-off. That is why analytic knowledge is self-reflection.⁴

Now this characterization of the therapeutic process is in some ways plausible and illuminating - though, as I will argue in the following section, it is associated with a view of the goal of psychoanalytic therapy that departs significantly from Freud's. But what is not so plausible is Habermas's further claim that the meaning of the concepts 'id' and 'ego' in Freud's metapsychology can and must be understood by reference to the therapeutic process characterized in this way. Habermas's formula here is simple: 'Id = alienated ego', as we can see in the following passage:

Freud conceived the defensive process as the reversal of reflection, that is, as the process, analogous to flight, through which the ego conceals itself from itself. *'Id' is then the name for the part of the self that is externalized through defense*, while 'ego' is the agency that fulfills

the task of reality-testing and censorship of instinctual impulses. [My italics]⁵

This definition of the id is supported by Habermas's translator, who tells us that:

Das Ich means the I and *das Es* the it. That is, they refer to the antithesis between reflexive, personal subjectivity and reified, impersonal objectivity.⁶

But this is not how Freud himself conceived of the id; or, at least, this is only a partial account which excludes what was equally important for Freud, namely the instincts (or their psychological representatives, a concept I will discuss shortly). This view is maintained by Freud throughout what is in this context a central work, *The Ego and the Id*, where we find several passages such as this:

It is easy to see that the ego is that part of the id which has been modified by direct influence of the external world through the medium of the Pcpt.-Cs [Perceptual-Conscious system]; in a sense it is an extension of the surface-differentiation. Moreover, the ego seeks to bring the influence of the external world to bear upon the id and its tendencies, and endeavours to substitute the reality principle for the pleasure principle which reigns unrestrictedly in the id. For the ego, perception plays the part which in the id falls to instinct. The ego represents what may be called reason and common sense, in contrast to the id, which contains the passions.⁷

And Freud goes on to say of the ego that:

in its relation to the id it is like a man on horseback, who has to hold in check the superior strength of the horse; with this difference, that the rider tries to do so with his own strength while the ego uses borrowed forces. The analogy may be carried a little further. Often a rider, if he is not to be parted from his horse, is obliged to guide it where it wants to go; so in the same way the ego is in the habit of transforming the id's will into action as if it were its own.⁸

This analogy is pursued later on, in a slightly different form, when Freud presents his well-known picture of the ego as:

a poor creature owing service to three masters and consequently menaced by three dangers: from the external world, from the libido of the id, and from the severity of the superego.⁹

However, whilst Freud's conception of the id departs in this respect from Habermas's, we should note that it does none the less include what Habermas views as the 'alienated ego', namely what has been repressed from consciousness during the kinds of conflictual situations, such as occur during the

Oedipal period, which are central to Freud's dynamic theory of psychosexual development and the formation of neuroses. In *The Ego and the Id*, for instance, having described how the ego may be seen as becoming differentiated from the id via its access to the external world through the perceptual-conscious system, Freud comments:

But the repressed merges into the id as well, and is merely a part of it. The repressed is only cut off sharply from the ego by the resistance of repression; it can communicate with the ego through the id.¹⁰

Thus, briefly, whilst for Habermas 'id = alienated ego (= the repressed)', for Freud 'id = the repressed + the instincts'. More fully, Freud's position is well described by Marie Jahoda in the following terms:

The id is a summary term comprising various functions, such as instinctual drives - hunger, thirst, sexuality - which mobilize the whole person and thus have overt consequences. These consequences are behavioural not just physiological, but the instinctual drives are, of course, 'open to the somatic', in Freud's terminology. There are also other functions subsumed in the id which, similarly, have mobilizing consequences but stem from previous experiences which become repressed; that is, there are memories of ideas, events, actions and feelings excluded from consciousness but activating the person.¹¹

Thus whereas for Habermas the apparent objectivity of the id is only apparent, since it is really an alienated form of subjectivity which can be restored to consciousness, this is so for Freud only of one component of the id, the repressed, and not for the other, the instincts. In effect, Habermas wishes to understand the id in the same way as, in the previous chapter, I suggested he understood the 'naturalness' of distorted communication, and the reified character of ideologies: their apparent objectivity is in fact an alienated subjectivity.¹² The id is not then an ineliminable feature of all human activity, as it was (partly) for Freud, but an element in the defective but alterable character of the object-domain of a critical social theory.

It might seem that some support for Habermas's view is given by Freud's well-known dictum, 'Where Id was, there Ego shall be' which Jeremy Shapiro, in one of his translator's notes to *Knowledge and Human Interests*, says should be taken to mean 'Where it-ness was, I-ness shall come into being'.¹³ But this ignores the important qualifications in Freud's own work to the meaning of this dictum, which will be explored further in the next section. For Freud, roughly, the id cannot be replaced by the ego, but can only be more effectively and less unhappily controlled or served by it. This is because he views the instincts as ineliminable; and although the specific 'objects' of their driving force may be

altered, the pressure of their demands for satisfaction cannot.

However, the contrast between Habermas's and Freud's accounts of the id is more complex than so far presented. This is partly because of the notorious difficulties in conceptualizing Freud's account of the instincts, particularly their biological nature;¹⁴ and partly because Habermas does not simply ignore their place in Freud's theory. Rather, he argues that to the extent that Freud conceived of the instincts in biological terms, he was guilty of a scientific misunderstanding of the epistemological character of his own theory (a misunderstanding evidenced also in his neurologically based energy-model of psychological processes, which I will discuss in the final section of this chapter). So we need now to examine more closely Freud's conception of the instincts.

Freud maintained that the concepts of psychoanalytic theory should be exclusively psychological: they must not be couched directly in physiological or biological terms.¹⁵ In the case of the instincts, this meant they could only be admitted into the theory in the form of what he often called their 'psychical representatives'. His terminology is not always consistent, but I do not think this affects the significant issues here. Sometimes he talked of the instincts *having* psychical representatives; at other times, of their *being* such representatives. The latter usage is involved, for instance, in the following two passages:

By an 'instinct' is provisionally to be understood the psychical representative of an endosomatic, continuously flowing source of stimulation The source of an instinct is a process of excitation occurring in an organ and the immediate aim of the instinct lies in the removal of this organic stimulus.¹⁶

And likewise he defines the 'source' of an instinct as:

the somatic process which occurs in an organ or part of the body whose stimulus is represented in mental life by an instinct. We do not know whether this process is invariably of a chemical nature or whether it may also correspond to the release of other, e.g. mechanical, forces. The study of the sources of instincts lies outside the scope of psychology.¹⁷

Elsewhere, however, Freud proposes a distinction between instincts and their ideational representatives, whilst accepting a 'looseness of phraseology' that may blur this. For instance:

An instinct can never become an object of consciousness - only the idea that represents the instinct can. Even in the unconscious, moreover, an instinct cannot be represented otherwise than by an idea When we nevertheless speak of an unconscious instinctual impulse or of

a repressed instinctual impulse, the looseness of phraseology is a harmless one. We can only mean an instinctual impulse the ideational representative of which is unconscious, for nothing else comes into consideration.¹⁸

In saying that these terminological variations are insignificant, I mean that whichever is used, Freud's basic claim remains the same: that the (instinctual) demands of the id are somatically based. As he says in the *New Introductory Lectures*, in discussing the instincts 'we are studying the psychical accompaniments of biological processes', and he goes on to comment:

We can distinguish an instinct's source, object, and aim. Its source is a state of excitation in the body, its aim is the removal of that excitation; on its path from its source to its aim the instinct becomes operative psychically.¹⁹

I do not think Freud ever wavered from this view, despite several (in other contexts very important) changes in his account of *which* instincts humans in fact have, and of their relationships to one another.²⁰ This view is perfectly compatible with his insistence that the concepts of psychoanalytic theory itself be exclusively psychological. And it is also consistent with three further, and equally strongly held convictions: that all psychological phenomena, and not just the instincts, have a somatic foundation; that psychoanalytic theory must be open, in principle, to a neurological reduction; and that psychological relationships are subject to an overall scientific determinism.²¹

Now Habermas opposes these reductionist and deterministic assumptions in Freud's work, and regards them as important elements in its scientific misunderstanding. In order to purge psychoanalytic theory of these assumptions, he thinks it necessary, amongst other things, to revise the way that Freud conceived of the instincts. Thus he criticizes Freud for proposing:

an energy model of instinctual dynamics with an objectivist turn. Thus he sees even the species' process of civilization as linked to a dynamic of the instincts. The libidinal and aggressive instinctual forces, the prehistorical forces of evolution, permeate the species subject and determine its history. But . . . [t]he conception of the instincts as the prime mover of history and of civilization as the result of their struggle forgets that we have only derived the concept of impulse privatively from language deformation and behavioral pathology. At the human level we never encounter any needs that are not already interpreted linguistically and symbolically affixed to potential actions.²²

A similar view is expressed in a more recent interview:

I think that the drives are something real in distorting verbal communication, they have some force in destroying an individual's capacity to act; but I cannot account for this without referring to a communication frame.²³

In both these passages, two main claims seem to be made. First, that in understanding human action the concept of an instinct is legitimate only if the character of instinctual impulses is specified in an interpretively intelligible form: that is, not biologically, but in terms of linguistic categories referring to possible human desires or actions. Second, that even these (linguistically interpreted) drives are relevant only in the explanation of distorted or pathological activity.

This latter claim involves a major departure from psychoanalytic theory, which I will explore in the next section. But considering now only the former claim, I suggest that it is not altogether inconsistent with Freud's own position. For, as I have noted, Freud himself insists that only the 'psychical representatives' of the instincts (or of their somatic biological sources) are appropriate elements within psychoanalytic theory; and this appears to meet Habermas's requirement of their being interpreted 'linguistically'. For instance, Freud says that instinctual drives are highly variable in the 'objects' to which they are directed at different times, as compared with what he calls their 'aim', which generally remains constant.²⁴ Presumably, drives specified in terms of their objects at particular times would satisfy Habermas's requirement, since these objects will typically be described in terms of linguistically defined actions, attitudes, desires, and so on. So this requirement would only involve a departure from Freud's position if it were also true that 'linguistically interpreted' instinctual impulses are not subject to causal determination, at the level both of psychological relationships, and of their (in Freud's view) biological bases.

It seems that the primary rationale for Habermas's rejection of biologically conceived instincts is his desire to remove the 'objects' of psychoanalytic theory from the domain of an empirical-analytic science which investigates causal relationships between non-meaningful items. But suppose instead, as I think is true, that there can also be causal relationships between meaningful objects. Then the determinism Habermas wishes to avoid by his insistence on a non-biological conception of instinctual impulses, will not thereby be eliminated. Freud, as I have noted, endorsed both a deterministic reductionism, and the use of exclusively psychological concepts in psychoanalytic theory. But Habermas believes that the hermeneutically interpretable realm of communicative activity is non-deterministic: that the category of causality is inapplicable to it. Clearly, this view is incompatible with Freud's deterministic reductionism. But the requirement that there be no biologically defined concepts in psychoanalytic theory does not by itself ensure that this is avoided. For it may be that both determinism and reductionism can be specified in ways that are consistent with this requirement, as I

will argue later in this chapter.

To summarize so far, Habermas's mis-interpretation of Freud's concept of the id expresses his wish to purge psychoanalytic theory of deterministic and reductionist assumptions. The missing element in his account of the id is the concept of the instincts. But Habermas does not ignore the instincts: instead, he tries to de-biologize them in a way that is, at first sight, in accordance with Freud's own practice of including within psychoanalytic theory only their 'psychical representatives'. Yet this by itself cannot satisfy Habermas, if it is in fact consistent with what he wants to reject, the deterministic reductionism of Freud's 'scientific misunderstanding' of psychoanalysis. And he thinks this has to be rejected because it is incompatible with the possibility of human autonomy, and with the characterization of psychoanalytic therapy as a self-reflective, emancipatory process. But this I believe is a mistake. Neurophysiological reductionism, properly understood, is compatible with a defensible conception of autonomy, and of psychoanalysis as a means to its realization.

However, we must now examine another of Habermas's departures from Freud, concerning the relationship between unconscious determinants and the goal of the psychoanalytic process. As I have pointed out, Habermas claims that instinctual impulses are relevant only to the explanation of defective activity. This non-Freudian view is related to his equally non-Freudian tendency to regard all unconscious determinants as distorting influences, that are to be eliminated in the self-reflective movement towards autonomy. As with his account of the id, this not only misinterprets Freud but also indicates defects in his conception of critical social theory.

2. The psychoanalytic goal: abolition of the unconscious or strengthening of the ego?

It will be helpful to begin by noting the relationship in Freud's theoretical work between the 'structural' model of id, ego and superego, and the 'topographical' model of the unconscious, preconscious and conscious. Whilst the former was introduced by Freud later on in his development, he did not intend it to directly replace the latter, since the two models were designed to meet different, but compatible, theoretical demands. Roughly speaking, the topographical model is aimed at distinguishing various kinds of psychological processes, and at identifying their distinctive modes of operation. Thus Freud argued that in addition to conscious processes (those of which we are aware at any time), there were not only preconscious ones (such as memories that could be brought to consciousness 'at will', as it were), but also unconscious ones that were far less immediately accessible, and could only be made conscious by overcoming resistances through various therapeutic techniques.²⁵ He went on to claim that the unconscious, and the conscious and preconscious 'systems' operated in quite different ways, described in his accounts of what he called the 'primary' and 'secondary' processes.²⁶ By contrast, the structural model is concerned with specifying the three main psychological functions that must be performed within the whole personality; the various sources of

conflict between these; and the possible modes of resolution of such conflicts. These functions are assigned, in an at least partly metaphorical way, to the three 'agencies' defined in the model. The difference of aim in constructing these models is reflected in the absence of any simple correspondence between their respective concepts: not even between the conscious (and preconscious) and the ego, since Freud notes that some of the ego's functions are performed unconsciously, particularly the various defence mechanisms, including repression.²⁷

Freud initially postulated the existence of unconscious mental processes in order to explain certain features of neurotic behaviour and of his early clinical experience in treating this (such as the use of hypnosis in cases of hysteria), whilst remaining at the level of a specifically psychological theory. But he later argued that important evidence of how these processes operated was revealed by the analysis of dreams. Dreams, he argued, were disguised wish-fulfillments - except in the case of (some) infant dreams, in which there was no disguise to be analytically penetrated; and he declared that 'the interpretation of dreams is the royal road to a knowledge of the unconscious activities of the mind.'²⁸ In deciphering the manifest content (the dream as reported) the analyst and patient in effect reverse the process of the dream-work, through which its latent content (the unconscious, repressed wish) has been disguised. Freud argued that there were various features of this dream-work that displayed the general mode of operation of the primary processes, such as condensation, displacement, the use of visual imagery and of symbols. The last of these, unlike the first three, involved meanings that were not peculiar to each patient: there was a (possibly universal) language of symbols, especially concerning the male and female genitals, used not only in dreams but in folklore, myth, and so on.²⁹

It is important to emphasize that, for Freud, dreams were not pathological phenomena, and neither were the dreams of neurotic patients essentially different from those of non-neurotic people. This is related to a more general claim: that there is nothing pathological, as such, about the presence and mode of operation of unconscious processes. He says, for instance:

we must recognize that the psychical mechanism employed by neuroses is not created by the impact of a pathological disturbance upon the mind but is present already in the normal structure of the mental apparatus.³⁰

And he suggests that the normal mind's activities are enriched by the fact that both unconscious and conscious (or preconscious) processes are involved:

For illnesses . . . do not presuppose the disintegration of the apparatus [the mind] or the production of fresh splits in its interior. They are to be explained on a dynamic basis - by the strengthening and weakening of the various components in the interplay of forces, so many of

whose effects are hidden from view while functions are normal. I hope to be able to show elsewhere how the compounding of the apparatus out of two agencies [the unconscious and the conscious: Freud had not yet introduced his structural model at this time] makes it possible for the normal mind too to function with greater delicacy than would be possible with only one of them.³¹

(I take it that Freud is referring here to the operation of the primary processes in fantasy, art, literature, jokes, and so on.)

Further, Freud says that in calling the processes of the unconscious ‘primary’ he is claiming that they have a chronological priority in individual development. The secondary processes of the conscious-preconscious emerge ‘only during the course of life’, and though they may eventually come to achieve control over the unconscious, the latter is never eliminated. Indeed:

In consequence of the belated appearance of the secondary processes, the core of our being, consisting of unconscious wishful impulses, remains inaccessible to the understanding and inhibition of the preconscious; the part played by the latter is restricted once and for all to directing along the most expedient paths the wishful impulses that arise from the unconscious. These unconscious wishes exercise a compelling force upon all later mental trends, a force which those trends are obliged to fall in with or which they may perhaps endeavour to divert and direct to higher aims.³²

More than twenty years later, in *The Ego and the Id*, he expresses a very similar view, but now articulated in the theoretical concepts of the later, structural model. Immediately following the passage I quoted in the first section, concerning the ‘three masters’ of the ego,³³ he goes on to say:

As a frontier-creature, the ego tries to mediate between the world and the id, to make the id pliable to the world, and by means of its muscular activity, to make the world fall in with the wishes of the id. In point of fact, it behaves like the physician during an analytic treatment: it offers itself, with the attention it pays to the real world, as a libidinal object to the id, and aims at attracting the id’s libido to itself. [I take it that Freud is here referring to the phenomenon of transference.] It is not only a helper to the id; it is also a submissive slave who courts his master’s love.³⁴

The picture presented here of the ego as the ‘submissive slave’ of the instinctual id suggests a parallel to Hume’s much-quoted comment that ‘reason is the slave of the passions’;³⁵ and it is associated with a conception of the overall goal of psychoanalysis on Freud’s part that differs in several important

ways from Habermas's. For Freud, I believe, this goal consists primarily in increasing the extent to which this 'slavery' can be successfully accomplished. The basic ends of human action are determined by the (innate) instinctual impulses of the id, governed by the pleasure principle. The task of the ego, governed by the reality principle, is to maximize the effective satisfaction of these impulses, and to minimize the suffering that results from various forms of their inhibition and repression. In performing this task, the ego is also hindered by the superego; and thus the realization of the goal of psycho-analysis also requires some reduction in the latter's tendency to generate avoidable suffering in the form of 'moral' guilt or anxiety.³⁶

Underlying this account is a version of ethical and psychological hedonism, in which the ego is the 'agency' of a prudential, instrumental rationality. This hedonism, however, becomes more complex with the introduction of the death instinct in Freud's later works, where it is argued that the aggressive re-direction of this instinct poses considerable problems for the maintenance of 'civilization' and thus, indirectly, for the amount of pleasure that individuals are able to experience. Freud claimed, amongst other things, that the self-punitive operations of the superego were to some extent necessary to counter the otherwise destructive consequences of these aggressive impulses; though he still believed that the degree of guilt and anxiety produced by the superego in modern societies was excessive, and could be reduced through therapeutic procedures.³⁷

Thus Freud's dictum, 'Where Id was there Ego shall be', can only be properly understood as subject to the qualifications implied by the above account. By contrast, Habermas's conception of the goal of psychoanalysis involves a quite different interpretation. Having (mis)-understood the concept of the id as the alienated ego, he presents in effect a literal and unqualified reading of this dictum, so that the abolition of the id is seen as a possible and desirable outcome of the therapeutic process. Likewise the instincts are regarded as the sources only of pathological, neurotic activity; and indeed the same is true of all unconscious determinants. Further, he maintains that in overcoming the power of the unconscious, the patient is thereby removed from the deterministic realm of causality; and on this basis he attempts to distinguish the emancipatory character of psychoanalysis (and thus, more generally, of a critical social theory) from the technical, instrumental character of empirical-analytic science:

In technical control over nature we get nature to work for us through our knowledge of causal connections. Analytic insight, however, affects the causality of the unconscious as such. Psychoanalytic therapy is not based, like somatic medicine, which is 'causal' in the narrower sense, on making use of known causal connections. Rather, it owes its efficacy to overcoming causal connections themselves.³⁸

And in contrasting what he calls the ‘depth-hermeneutic understanding’ (a concept I will examine in the next section) involved in psychoanalysis with the causal explanation of empirical-analytic science, he says:

Depth-hermeneutic understanding takes over the function of explanation. It proves its explanatory power in self-reflection, in which an objectivation that is both understood and explained is overcome. This is the critical accomplishment of what Hegel had called comprehending (*Begreifen*).³⁹

The reference here to Hegel is worth exploring briefly, for it recalls Habermas’s comments on Marx’s materialism discussed in section 3 of the previous chapter. Habermas endorses Marx’s refusal to accept Hegel’s transcendence of the subject-object duality, where nature is the ‘object’, and humans the ‘subject’. Thus, whereas for Hegel all apparent objectivity can be transcended, being regarded as an alienation of *geist*, this is not so for Habermas. It is only so in the case of the apparent objectivity of alienated, reified human subjectivity, of which the id and the unconscious are for him central examples. And it is alienated subjectivity that forms the object-domain of critical social theory, which aims to overcome this domain’s alienated character in the movement towards human emancipation. Psychoanalysis then, for Habermas, aims to do just this. It is therefore not to be understood as empirical-analytic science, directed towards technical control through knowledge of causal relationships, but as a critical science directed at the abolition of what only appear to be genuinely causal determinants. Yet, as my account of Freud has been intended to show, this clearly departs from Freud’s view, which comes close to specifying the goal of therapy as increasing the ‘technical’ control exercisable by the ego over both the instinctual impulses and the external world, both being equally ‘objective’. There is, as far as I can see, no hint in Freud’s position of any opposition to the kind of instrumental conception of reason which Habermas, like other critical theorists, wishes to replace or extend by an alternative view of reason.

Finally, through his association of the de-alienation of the ego with the ending of unconscious determination, Habermas seems to commit himself to the eliminability of all unconscious influences on human activity, and thus of their distinctive mode of operation, the primary processes. We can bring out the possible significance of this by asking: does Habermas then believe that a goal of therapy is the abolition of dreams? For according to Freud dreaming is a non-pathological phenomenon whose character reveals the nature and primacy of the unconscious. Further, it is not only dreams but many other non-pathological activities that display these primary processes which involve, as say in fantasy, the use of condensation, displacement, visual imagery and symbols. So we can ask also: does Habermas regard these, too, as pathological? I do not see how he can avoid this

conclusion, without a major revision of his reinterpretation of psychoanalytic theory.⁴⁰

However, as I noted at the outset of this chapter, one cannot directly reject Habermas's conception of critical social theory merely by showing that it does not 'fit' the actual character of the theory he regards as providing an epistemological model for it. And in any case Habermas himself argues that there are elements of a scientific misunderstanding in Freud's work, which he wishes to remove from it - though, at the very least, it appears that Habermas does not fully recognize the extent to which his preferred (re)interpretation of psychoanalysis differs from Freud. But what matters most is the legitimacy of the assumptions that lead Habermas to make these major revisions; and it is to this issue that the rest of the chapter is addressed. In particular we need to examine his attitude towards determinism and reductionism, and the reasons why he apparently regards these as incompatible with the understanding of psychoanalysis as emancipatory. I shall argue that he is mistaken in thinking this, starting with the question of determinism and human autonomy.

3. Autonomy, causality and compulsiveness

Underlying Habermas's view of the way in which psychoanalysis is to be seen as emancipatory are certain assumptions about the relationship between autonomy and causality. To identify these it will be helpful to review briefly the debates about free-will and determinism, and about the status of purposive explanation, that have taken place within analytical philosophy.

On the former question, the basic disagreement has been between those who regard freedom and determinism (defined as the thesis that all events are fully causally determined) as compatible, and those who do not. Compatibilists have argued that the distinction between free and unfree actions lies not in the respective presence or absence of causal determinants, but in differences in the nature of these causes. By contrast, those who reject compatibilism have opted either to accept determinism and deny the existence of freedom (so-called 'hard determinists'), or conversely to assert the existence of freedom, and deny the truth of determinism (the 'libertarians').⁴¹

In the partly related debates about the nature of purposive explanation (that is, explanations of human action that make reference to the agent's intentions, aims, motives, and so on), the main issue has been whether this can be regarded as a form of causal explanation. Here, the problem has centred around the generally (though not universally) accepted Humean criterion that, for two items to stand in a causal relationship with one another, there must be a possible identifying description of each that is logically independent of the other. That is, there must be a referential description for each item, such that the truth or falsity of existential claims made using this description, for either item, does not logically entail the truth or falsity of an existential claim about the other. Thus it has been generally agreed that if there is no way of identifying an agent's intentions that is logically independent of his or

her performance of the action said to be explained by reference to this intention, then the nature of this explanation cannot be causal. Further, amongst supporters of the causal status of purposive explanation, some have argued that the requisite independent identification of the agent's purpose can be provided within the conceptual framework of purposive language; whilst others have argued that this identification can only be achieved by moving to a neurophysiological level. The latter view has normally been associated with the adoption of an identity-theory concerning the relationship of the 'mental' and the 'physical', according to which the former concepts are taken to in fact refer to physical items.⁴²

There is a complex set of relationships between the various positions within each of these debates. I will note only two of these. First, libertarianism (at least, those versions of it which take purposive action as a paradigmatic case of free action) requires rejection of the causal analysis of purposive explanations. Second, both compatibilism and hard determinism are at least consistent with this causal analysis, since both accept the possibility of causal explanation for all phenomena, including human actions.

I shall not argue the respective merits of various positions within these debates. But I will assume the causal status of purposive explanation, accepting the analysis of this as involving desires and beliefs as causal determinants of action; and I shall assume also that the logical independence requirement can be met without recourse to neurophysiological identifications.⁴³ On the question of determinism I shall present a version of compatibilism, and try to show how Habermas's conception of psychoanalysis as emancipatory may be reconstructed from this standpoint, together with the view of purposive explanation just noted. However, it should be noted that I am not thereby committed to the view that compatibilism provides the correct solution to all issues about free-will and determinism. There may be senses or forms of freedom that are incompatible with determinism. My suggestion is only that we can understand the idea of psychoanalysis as an emancipatory process by using a conception of autonomy that accords with a compatibilist standpoint.⁴⁴

Habermas, it appears, maintains a non-causal analysis of purposive explanation, together with a form of libertarianism. But he does not regard all human activity as (already) free: autonomy is an ideal from which much past and present human action falls short. Unfree action is alienated, reified action that remains subject to a specific form of causal determination, which he calls the 'causality of fate', or 'second nature'.⁴⁵ And, at least in the case of neurosis, it is the unconscious, or the id (understood in his terms) that provides this freedom removing causality.

Habermas's denial of the causal status of purposive explanation is displayed in his view of hermeneutic interpretation, and its contrast with what he calls the 'depth hermeneutics' of Freudian

theory. He claims that the kind of communicative activity explored in psychoanalysis is not open to successful interpretation by means of normal hermeneutic procedures. The text of the manifest dream content, for instance, does not contain occasional slips or missing items, as might a written text studied by philologists, but what he describes as ‘systematic’ distortions. Their interpretation requires the employment of a ‘depth hermeneutics’, which is said to explain these distortions by ‘the psychological investigation of causal connections’.⁴⁶ By contrast, ‘ordinary’ hermeneutic interpretation is concerned only with the understanding of meanings; and the identification and understanding of actions by reference to the agent’s purpose or intentions are included within this form of interpretation. That is, ordinary hermeneutic interpretation includes the understanding of actions gained through their intentional descriptions.⁴⁷ Thus the contrast between this, and depth hermeneutics, is specified at least partly by the presence of causal explanation in the latter, and its absence in the former.

However, for Habermas this is not the only contrast between the two types of hermeneutic interpretation. For in the interpretation of dreams (and thus also, of neuroses), a crucial feature is the discovery of the transformations of the latent content, the unconscious wish, by the mechanisms of the dream-work: condensation, displacement, and so on. The discovery of these transformations enables us to decipher the initially unintelligible text. Whilst remaining within the essentially hermeneutic concern with meanings, normal interpretive procedures are in this way significantly extended.

I now wish to revise Habermas’s position on this issue by proposing that the contrast between these two kinds of interpretation be made on the basis of the second contrast alone. That is, what makes depth hermeneutics ‘deep’ is only the distinctive character of its interpretive procedures, and not (also) its investigation of causal relationships. This would have the virtue (for me) of being consistent with accepting a causal analysis of purposive explanation, whether the purposes are conscious or unconscious. In both cases, I suggest, we are necessarily involved in the interpretation of meanings; but we are also necessarily concerned with questions of causality. We are involved in the former, roughly, because the specification of purposes or intentions involves the identification of beliefs and desires; and the identity of these is given in terms of their propositional content. It is this content that requires interpretive understanding. But whether the agent’s beliefs and desires (and, indeed, which of them) are the causal determinants of his or her actions, is a distinct, and not interpretively decidable, matter. On this view, then, any social science must be concerned with both interpretive and causal questions at every level of analysis, whether the ascertained purposes are conscious or unconscious.

This, then, is the first revision of Habermas’s position that I propose. The second concerns the analysis of freedom and determinism. Here I suggest a form of compatibilism, which preserves a good deal of the way he wishes to contrast autonomy with its absence, but which is consistent both with

determinism, and with the form of reductionism I will present in the next section. There is, of course, an important link between compatibilism and my first revision: to regard purposive explanation as causal requires, other things being equal, that one either rejects the existence of freedom altogether, or regards it as compatible with determinism.

Compatibilists, as I noted earlier, typically distinguish free and unfree actions on the basis of the kinds of causal determinants that are operative. But the criteria for differentiation vary, in different forms of compatibilism. One criterion that has been proposed is the presence or absence of coercion; and a distinction is often drawn between ‘external’ coercion, for instance threats or sanctions, and ‘internal’ coercion. It is the latter which is, I think, the more fruitful to pursue, in order to understand how the concepts of autonomy and emancipation, especially in psychoanalysis, may best be analysed. However, I doubt if the concept of coercion is really appropriate here: a better one is ‘compulsiveness’.

We can begin to analyse this concept by considering some examples of activities typically regarded as neurotic, such as compulsive hand-washing or theft. In what sense are these compulsive? I propose something along the following lines. In compulsive actions, agents find themselves unable to act in accordance with their beliefs and preferences. For instance, in compulsive hand-washing, the agent typically suffers considerably from the consequences of persistent washing, and sees no benefit accruing from this activity that, in terms of his or her other desires, outweighs this suffering. Yet no amount of attention to the undesired character of these consequences, no amount of effort or determination to disengage from the activity, has any effect. The activity continues, despite constant attempts to end it: the normal degree of control over activity exercised by one’s beliefs and desires is absent. A similar situation obtains with kleptomania: despite believing that no desired consequences will follow from theft, the agent is unable to control activity to accord with this attitude.

It should be noted that, on this analysis, there can of course be compulsive *failure* to steal, just as well as compulsive stealing. This would be in line with Freud’s view of the operation of the superego as, in effect, a compulsive determinant of ‘moral’ behaviour. Thus Wilhelm Reich, in his attacks upon ‘morality’, basically criticized the compulsive character of superego-sustained ‘moral’ behaviour, and advocated instead the ideal of ‘self-regulation’. Freud, by contrast, came to believe that a certain degree of compulsive morality, enforced by the superego, was necessary for the existence of what he termed ‘civilization’, and thus also for the happiness of individuals. However, the disagreement between Reich and Freud turns largely upon whether Freud was right about the existence of the death instinct, which Reich denied.⁴⁸

But this initial sketch of the concept of compulsiveness needs to be elaborated to meet an important objection: is it not possible for activity to be compulsive, despite the fact that it *does* conform to the agent's beliefs and preferences? For instance, my smoking cigarettes might be compulsive, despite my believing that on balance the pleasure it provides outweighs the risks of suffering involved, and having no other reasons for regarding it as undesirable. To deal with such examples we need to introduce a hypothetical (counterfactual) element into the analysis. Then, smoking would count as compulsive if it were true that, *even if* I came to believe that the risks outweighed the benefits, or indeed that it was a thoroughly undesirable activity that I no longer had any wish to continue, I would continue to find myself doing it.

However, the need for another qualification - and one that indicates certain difficulties in specifying an adequate concept of autonomy - is suggested by further consideration of the same example. Perhaps, although it is true that if my beliefs and desires changed, the activity would cease, it is none the less compulsive because, in fact, I would display some kind of 'compulsive' resistance to any attempts made to alter these beliefs and desires. For instance, I might stubbornly refuse to take note of medical evidence concerning the risks of smoking; ignore claims made by companions about the discomfort or disgust they experience in my smoke-generating presence; and so on. I display, in some sense, a compulsive impenetrability to relevant information and argument, and remain unmoved by good reasons.

If the analysis of compulsiveness (and thus of autonomy) is to be adjusted to deal with this problem, it seems that we have to talk not just of compulsive actions but also of compulsive, or compulsively held, reasons for actions. In other words, the concept of compulsiveness must include a certain kind of 'resistance to reason'. But this is difficult to define. First, it would seem to require a general specification of what rationality consists in; and this would have to include many different kinds of reason, or reasoning about many different kinds of things - scientific beliefs, moral and political values, the relevance of feelings or emotions to action, and so on. Second, it would be necessary to provide criteria for discriminating between mere disagreement with, and the compulsive rejection of, what is rational (in terms of these standards). I have no suggestions to make here on either of the points. But I think the basic character of this analysis of compulsiveness is correct, and will continue to use it in what follows.

To show how this account can be used to reconstruct Habermas's view of psychoanalysis as emancipatory, we need now to relate the concept of compulsiveness to Freud's theory of the neuroses. This can be done, at least in general outline, by suggesting that the central feature of neurotic activity is its compulsive character; and that this, according to psychoanalytic theory, is due to its control by the repressed elements of the unconscious. The therapeutic process, then, involves the identification

and overcoming of those unconscious determinants that render behaviour compulsive. This is what the self-reflective movement from non-freedom to autonomy achieves. It is not, though, a movement from causality to the absence of causality: autonomous activity is causally determined, but in a non-compulsive manner. Thus the therapeutic process is to be seen as involving the replacement of one kind of causal determination by another. Precisely how this is achieved is a question of the dynamics of the therapeutic situation, in which various techniques are employed by the analyst to enable the transformation to occur.

Further, on the view I am suggesting here, we must also regard the character and outcome of the therapeutic process as causally explicable; though, as will be argued in the next chapter, a theory that explains this process may be only loosely related to a theory that explains the formation of compulsive activity itself. This deterministic view of the therapeutic process does not deny the possible significance, for instance, of self-reflective engagement by the patient as a primary means by which compulsive determinants are removed. But it does require that we see the capacity for self-reflection, its actual operation, and the changes that it can bring about, as causally explicable. In the following sections, I will present a version of neurophysiological reductionism which is, I think, consistent with this standpoint. But before that, some further comments on how this analysis of compulsive activity and the goal of psychoanalytic therapy differs from Habermas's position need to be made.

I argued that Habermas's account of this goal as, in effect, the abolition of the unconscious, involved a departure from Freud in at least two ways: the place of the primary processes, and of the instincts, as ineliminable features of non-pathological activity. In relating the concept of the compulsive to psychoanalysis, I have referred only to the repressed unconscious, thereby excluding the elimination of either of these as a requirement for non-compulsive activity. It seems to me implausible to view phenomena such as dreams or fantasy as compulsive, as undermining autonomy.

As for the (biologically determined) instincts, their relationship to the concept of autonomy is more complex. If there are such instinctual drives then they must be seen at least as placing limits on possible forms of human activity, and thus upon the practical significance of certain possible normative ideals. That is, humans may come to have desires that are not in fact realizable, since their realization would be incompatible with instinctually determined limits. Of course, this issue is extremely problematic because of the difficulties in defining adequately the concept of instinctual drives.⁴⁹ (Nor, incidentally, does neurophysiological reductionism entail the existence of such items, as will be seen later.) But if there are such things as instinctual drives (or other such innate elements), it must follow that human autonomy has definite limitations. For it would mean that whatever we wanted or believed, there would be some activities that we would be unable to perform. Whether such

inabilities should be described as compulsive, is a somewhat different issue. For one might choose to restrict the application of the concept of the compulsive to those incapacities that are eliminable.

It was Freud's view of the instincts that partly led him, as I argued in the previous section, to regard the rationality of the ego as essentially instrumental, concerned to achieve the most effective satisfaction of their demands. We might wish to modify this picture, for instance by claiming that not all human desires can be understood as derivatives of instinctual impulses.⁵⁰ But unless we reject altogether the idea of biologically determined limits upon human activity, we cannot view the ideal of human autonomy in the Promethean way that some of its advocates often seem to, and which is perhaps a consequence of any philosophical position that separates radically the human species from the rest of nature, defining the former only in terms of its species-distinctive features, a view which I criticized in section 4 of the previous chapter.

4. The reduction of psychology to neurophysiology

Having proposed this compatibilist revision of Habermas's view of autonomy as the goal of psychoanalytic theory, I will now try to specify a version of reductionism which is consistent with the proposed concept of autonomy but which preserves an important element in what he regards as Freud's scientific misunderstanding of psychoanalytic theory, namely the possibility of neurophysiological reduction. As I suggested in the first section, a major reason for Habermas's rejection of Freud's biological instinctualism - and thus for his peculiar understanding of the id - is his general opposition to reductionism, an opposition which is related to his ontological dichotomy between humans and nature. I will argue that reductionism does not entail, though it is consistent with, Freud's instinctualism. Further, the form of reductionism I present does not justify the way that Freud based his conception of psychological processes on a (speculative) neurological model of energy-distribution. Like Habermas, I am sceptical about Freud's use of this model; but for rather different reasons, which I will describe in the final section of the chapter.

The best overall account of scientific reduction has been provided by Ernest Nagel, and I will follow his general approach, though with some modifications.⁵¹ Nagel claims that we should consider reduction as the derivation (and thus explanation, assuming a deductive-nomological model of this⁵²) of one scientific theory from another, where each theory is specified in terms of its basic laws. He points out that for such a derivation to be valid, the premisses must include not only statements of the laws of the 'reducing' theory, but also a further set of statements which link the concepts of the 'reduced' to the 'reducing' theory. This requirement reflects the general point that, in deductive arguments, there cannot (non-trivially) be concepts in the conclusion that do not occur in the premisses. These additional premisses he calls 'bridging statements', an example of which, in the case of the reduction of thermodynamics to statistical mechanics, would be the statement relating

temperature to the mean kinetic energy of molecules.

Nagel suggests that different forms of reduction can be analysed in terms of the different character of these bridging statements. Partly now diverging from his account, I suggest there are three possibilities. First, the statements may be analytically true, by virtue either of an existing identity of meaning between their respective concepts, or of definitional decisions. Second, they may express claims of contingent identity: that is, the statements' referential terms may be held to identify the same items. In this case, the apparent ontology of the reduced theory is in effect replaced by that of the reducing theory. Third, the bridging statements may specify regular but logically contingent relationships between the items referred to in the basic laws of the theories concerned, these relationships being of a possibly causal nature. I will call the three forms of reduction involving these different kinds of bridging statements, 'conceptual', 'ontological' and 'explanatory', respectively. (It is worth noting that the two best-known recent versions of philosophical materialism as a solution to 'the mind-body problem', namely logical behaviourism and the identity-theory, involve, respectively, conceptual and ontological reductions of mentalistic to physicalistic theories.⁵³)

It is often thought that any form of reductionism must involve denying the reality of what is 'reduced' to another theoretical level. And this itself is then thought to have undesirable normative implications where what is reduced includes various features of human activity, such as beliefs, desires, feelings, deliberation, and so on, to which considerable significance is normally ascribed in moral and political issues. But although this might be argued with some degree of plausibility in the case of ontological reduction, and perhaps also for conceptual reduction, I do not see how this can be true of explanatory reduction. For here it is not claimed that the items described by the reduced theory are 'really' no different from the items described by the reducing theory. Thus, to propose an explanatory reduction of psychology to neurophysiology does not entail that psychological phenomena are not 'real', and that only neurophysiological events, states, or processes exist. Indeed, quite the opposite seems true: explanatory reduction requires that the theories are about non-identical domains.

However, in saying this I depart from an important feature of Nagel's account. He insists that we should regard reduction as a relationship between (sets of) statements, and not between things or their properties.⁵⁴ So, whilst Nagel would in a sense agree that an explanatory reduction does not deny the reality of the reduced items, this is because reduction is concerned only with statements, and not with 'what they are about'. This claim about 'what is reduced' is, I believe, based on Nagel's arguments elsewhere in *The Structure of Science* about the merits of realism as an account of the cognitive status of theories. He says that the difference between realist and non-realist (instrumentalist) views is insignificant: it consists only in a 'conflict over preferred modes of speech'.⁵⁵ Now I do not accept this, and I also think that realism is the correct view of the cognitive status of scientific theories. I

cannot argue for these claims here; but I will continue to talk in the realist mode, which followers of Nagel can interpret as merely a mode of speech, and which anti-realists will find unpalatable: they will have to provide their own translation.⁵⁶

Where reductionism concerns the relationship between psychology and neurophysiology, the question of realism is of considerable significance (if Nagel is wrong). For explanatory reductionism, from a realist standpoint, involves commitment to something very like the traditional epiphenomenalist solution to the mind-body problem. If psychology is explanatorily reducible to neurophysiology, and both these sciences are about real items, it follows that psychological phenomena such as perceptual experience, bodily sensations, thoughts, feelings, and so on, are the outcome of neurophysiological events, states, or processes. That is, they are epiphenomenal, at least in relation to neurophysiological phenomena, though it would still presumably be possible that they operate as the immediate causes of behaviour, whilst always being effects in relation to neurophysiological items.

Now some people regard this consequence, epiphenomenalism, as showing that either explanatory reductionism or its realist interpretation must be mistaken, for epiphenomenalism has been generally judged as an obviously unacceptable view. But I do not find the usual arguments against it convincing.⁵⁷ One is that epiphenomenalism entails the paradoxical consequence that ‘the whole of human history would have been just the same even if there had been no psychological phenomena’. Actually, this does not follow. For if psychological phenomena are neurophysiologically determined then, had the former not occurred, then neither would the latter; and ‘human history’ would indeed have been very different. (This is because the occurrence of an effect is a necessary condition for the occurrence of its causes, where the latter are sufficient conditions of the former.) Another standard objection comes from an evolutionary perspective: if species-characteristics emerge because of their adaptiveness for the organism why, for instance, should humans experience pain if this has no effects, and cannot therefore be adaptive? But there is a possible answer to this: what is adaptive may be the neurophysiological bases of these sensations, and not the sensations themselves.

The most serious difficulties faced by epiphenomenalism concern the supposed unintelligibility of (even one-way) causal relationships between the ontologically distinct domains of the neurophysiological and the psychological. Perhaps this is an overriding objection. But two points may at least weaken its *prima facie* force. First, of the two traditional rivals to epiphenomenalism, interactionism and materialism, the former shares with epiphenomenalism the supposed unintelligibility of physical-mental causal relations, whilst the latter has its own obvious difficulty, namely its denial of the ontologically distinctive character of phenomena such as thinking and the experiencing of sensations and feelings. Second, I am not convinced that the ‘unintelligibility’ of neurophysiological-psychological causal relations is any greater than is involved in any form of inter-

level explanatory reduction in science. To parody the usual objection to epiphenomenalism: how *could* a change in the chemical composition of the receptor areas in synaptic transmission affect the electrical activity of the neuron?⁵⁸

After this brief excursion into the traditional mind-body problem, let us return to reductionism. Though I cannot argue this here, I believe it is reasonable to accept the possibility of an explanatory reduction of psychology to neurophysiology. At the very least, I think that this kind of reductionism is far more plausible than either of its alternative forms, conceptual or ontological reduction. And I now wish to consider the consequences of explanatory reductionism for the idea of psychoanalysis as an emancipatory process: arguing, in fact, that the two are quite compatible. This will mainly consist in showing that certain supposed consequences of reduction do not follow. Since reductionism has often been opposed on the grounds of its having these consequences, I take it that showing this not to be so provides indirect support for the position.

First, an explanatory neurophysiological reduction of psychology (which I will now call just ‘reductionism’) does not entail, though it is consistent with, innatism - that is, the view that certain features of the human species are in some sense genetically inherited. The truth or falsity of this latter view would, from a reductionist standpoint, depend upon the actual character of the ontogenetic development of the human neurophysiological system: in particular, upon the relationship between genotypic and environmental determinants. This is a highly complex issue, both conceptually and substantially. It is important, for instance, to recognize that it is conceptually incoherent to think of ‘innate’ characteristics as those that are determined by genetic factors *alone*, since, from the time of conception, the development of any of an organism’s characteristics must result from the relationship between genotypic and environmental features. None the less, it may be possible to distinguish between those characteristics that are (relatively) environment-invariant and those that are (relatively) not so: that is, between characteristics which will develop in more or less any of the ‘environments’ which members of the species normally encounter, and those that are far more sensitive to particular environmental variations. If the concept of innate characteristics is to be used at all, it must be understood in terms of this kind of contrast.⁵⁹ But whether or not there are, for the human species, characteristics that are in this sense ‘innate’ - and if so, which - is not a question to which reductionism entails a particular substantive answer.

Second, reductionism is compatible with the occurrence of what are often termed ‘psychosomatic’ phenomena, provided that their usual conceptualization is partly revised. These phenomena range from the relatively straightforward, such as the apparent generation of peptic ulcers by chronic anxiety conditions, to the more esoteric, such as typical somatic forms of hysteria. I will comment briefly on the latter, since the study of hysteria provided an important impetus for Freud’s introduction of the

concept of unconscious psychological processes. Hysterical symptoms, such as partial paralysis or anaesthesia, often displayed a lack of correspondence between the regions of somatic malfunction, and theoretically intelligible anatomical regions. Thus the malfunction could apparently not be explained by reference to disturbances of normal physiological processes. In other cases, whilst this obvious difficulty for orthodox medical explanation was absent, Freud none the less believed that a psychological explanation was appropriate. For instance, in the clinical histories presented in *Studies On Hysteria*, he and Breuer discuss the case of Frau Cäcilie M., who suffered from:

a violent pain in her right heel - a shooting pain at every step she took, which made walking impossible. Analysis led us in connection with this to a time when the patient had been in a sanatorium abroad. She had spent a week in bed and was going to be taken down to the common dining-room for the first time by the house physician. The pain came on at the moment when she took his arm to leave the room with him; it disappeared during the reproduction of the scene, when the patient told me she had been afraid at the time that she might not 'find herself on a right footing' with these strangers.⁶⁰

Frau Cäcilie seemed to them a veritable bundle of such afflictions. A penetrating pain between the eyes in her forehead was, according to her analysts, generated by what she had felt as a 'piercing' look from her grandmother; a stabbing sensation in the heart meant 'it stabbed me to the heart'; a pain, as of nails driven into her head, was analysed in relation to thinking 'Something's come into my head'; and an hysterical aura in her throat, occurring after being insulted, was interpreted via the thought of 'I shall have to swallow this'.

I find these psychoanalytic explanations in principle perfectly plausible - which is not to deny difficulties in their validation. Freud believed such phenomena could be understood by the postulation of unconscious mental processes; and his later development of the concept of the primary processes, in the context of his theory of dream-interpretation, enables one to see how somatic symptoms may be meaningfully related to unconscious thoughts and wishes. In the examples given above, we see the predilection of the unconscious for puns and metaphors, generating the appropriate bodily behaviour and sensations. But there is nothing here that is inconsistent with (Freud's own) commitment to neurophysiological reductionism, provided two assumptions are made: that the operation of the unconscious is itself neurophysiologically explicable; and that the relevant neurophysiological processes are causally related to others, involved more directly in the generation of the somatic hysterical symptoms.

Thus, strictly speaking, 'psychosomatic' phenomena are mis-named, in that their explanation must ultimately itself be 'somatic'. But the concept is none the less very useful. For, in terms of existing

knowledge about physical processes - such as the sources of muscular malfunction - there may be no possibility of explanations other than in psychological terms, even though if reductionism is correct, these are themselves to be explained by reference to the complex and largely unknown character of the central nervous system. (Incidentally, many psychosomatic phenomena involve malfunctions of the peripheral areas of the so-called autonomic nervous system that have no apparent causes locatable within that system. The causes are thus dubbed 'psychological'. My version of reductionism suggests that these causes may typically lie in the central areas of the so-called voluntary nervous system. This requires interactions between the two systems, which had until relatively recently been regarded as independent of each other. A defect of Reich's otherwise admirable emphasis on the significance of the somatic character of neurosis, was his exclusive concentration on the autonomic nervous system. This partly prevented him from recognizing the importance of the unconscious primary processes, whose neurological bases cannot be located there, in generating the somatic characteristics of neuroses.⁶¹)

To conclude this discussion of psychosomatic phenomena, it is worth noting one more example: that words can hurt. Clearly the acoustic input of, say, quietly spoken personal insults cannot explain the often painful sensations they cause. Nor indeed can the insulting character of the utterance normally be understood outside the context of the history of the relationship between insulter and insulted. Some highly complex psychological relationships appear to link the utterance to the sensation. Explanatory reductionism requires us to believe that this mediation is neurophysiologically based, and I see no reason why this should not be so.

I turn now to another supposed consequence of reductionism: that if reductionism is true, then psycho-therapy is in principle replaceable by 'neuro-therapy', that is, by some kind of physical intervention in the neurophysiological system. This, I think, is so; but the 'in principle' must be understood here in such a way that, 'in practice', this possibility is almost certainly unrealizable. (Were it to be realizable, however, there would still be important normative issues in the choice between psyócho- and neuro-therapy.) For the practical difficulties are not in any simple sense due to the relatively undeveloped state of neurophysiological science. The problem is not that little is yet known that would make this kind of replacement of therapeutic techniques possible, so that all that is needed is a lot more knowledge about a uniquely complex entity, the human central nervous system. Rather the problem is that what is already known about this system suggests that the kind of detailed reduction that would be necessary is probably unachievable.

I have in mind the following sorts of considerations. First, with some partial exceptions, the neurophysiological bases of different kinds of human activity are not conveniently located in spatially distinct and independently operating parts of the brain. And even in cases where they to some extent

are, in normal functioning, the brain generally appears to display a considerable degree of 'equipotentiality': that is, the same function can often be performed by different neuronal structures when the need arises.⁶²

Second, even in some of the most successful areas of neurophysiological research, such as investigation of the neuronal structures and processing mechanisms involved in visual perception, what has already been discovered suggests that any hope of a neurophysiological account that would be able to explain the particular perceptual experiences of individuals is unfounded. It is not merely that such experiences cannot be understood as simple functions of the visual information input (or of this together with other sensory information). The difficulty is that although human visual processing may involve mechanisms common to the species, the way in which these actually operate, in one individual at a particular time, will be affected by the past experience (both perceptual and non-perceptual) of that person, by his or her linguistic concepts, by present expectations and mood, and so on.⁶³ And whilst the general principles upon which each of these kinds of determinant operate may be discoverable, the idea of combining knowledge of these with information about the often unique features of each individual, or even groups of similar individuals, so as to predict or explain their perceptual experiences, is quite implausible.

One other example, from a different (and apparently much simpler) area of neurophysiological research, may be helpful: the study of eating and drinking. To some extent it has appeared possible to locate specific areas of the hypothalamus whose activity controls this behaviour, and to discover feedback systems for these which, for instance, may monitor the concentration of glucose in the bloodstream.⁶⁴ But there is obviously a vast gap between knowing this, and being in a position to explain or predict actual human eating or drinking. Past experience, cultural values, emotional states, beliefs about diet, and so on, all influence this. Eating and drinking are complex, socially and individually differentiated activities, and the belief (which I accept) that they have a neurophysiological explanation should not lead to any 'practical' optimism about the possibility of a detailed realization of reduction. For most practical purposes, psychological theories will continue to be far more useful.

What I have suggested here is very much in agreement with Margaret Boden's assessment of the prospects for the neurophysiological reduction of psychology:

the human neurophysiologist may never be able to offer more than the general principles of brain function. These may perhaps include some differential features of the cerebral mechanisms involved in broadly different types of purposive activity, such as the various instincts and other features marked by the key concepts in a personality theory. But the

physiologist may never be able to specify sufficient causal conditions of all those classes of human behaviour that are interesting enough to cry out for explanation and prediction. One could not then explain a man's threat to commit suicide, in any language, nor predict his wife's behaviour should he do so, merely by specifying their physiologically identified brain-states. One would have to rely - as now - on intensional expressions referring to their hopes, beliefs, desires, self-image, and general world-view.⁶⁵

And she concludes:

Psychology may be empirically reducible to physiology, in principle, even though contingent facts (such as cerebral equipotentiality, as well as the differences between natural languages and between social institutions and conventions) make it extremely unlikely that the reduction will be effected in anything but a highly schematic form. In other words, psychology may sensibly be said to be 'empirically reducible' to physiology in a weak sense that implies the total dependence of teleological phenomena on physical causal mechanisms, the absolute necessity of the mechanistic embodiment of the mind, but which does not imply that bridging statements will actually be found to correlate every psychological statement with a specific neurophysiological statement.⁶⁶

I introduced this discussion of what, in practice, were the prospects for reductionism in response to the objection that such a view would involve commitment to the replacement of psycho- by neuro-therapy. I turn now to consider the implications of what I have said for Habermas's objections to Freud's reductionism, which was at times associated with a belief in the possibility of this kind of therapeutic replacement.

5. Freud's neurological energy model

As Habermas notes, Freud viewed psychoanalytic theory as a strictly empirical science, and believed that it would eventually be reducible to neurophysiology.⁶⁷ Further, he sometimes claimed that it might one day prove possible to replace psychoanalytic therapy by neurophysiological techniques. For instance in his last work he said:

The future may teach us to exercise a direct influence, by means of particular chemical substances, on the amounts of energy and their distribution in the mental apparatus But for the moment we have nothing better at our disposal than the techniques of psychoanalysis.⁶⁸

And twenty years earlier, in the *Introductory Lectures*, discussing whether psycho-analysis was ‘a causal therapy’ (one that operated upon the causes, rather than the symptoms, of neurosis) he had written:

In so far as analytic therapy does not make it its first task to remove the symptoms, it is behaving like a causal therapy. In another respect, you may say, it is not. For we long ago traced the causal chain back through the repressions to the instinctual dispositions, their relative intensities in the constitution and the deviations in the course of their development. Supposing, now, that it was possible, by some chemical means, perhaps, to interfere in this mechanism, to increase or diminish the quantity of libido present at a given time or to strengthen one instinct at the cost of another - this then would be a causal therapy in the true sense of the word, for which our analysis would have carried out the indispensable preliminary work of reconnaissance.⁶⁹

Habermas argues that Freud’s reductionism, based on a model of the economics of energy-distribution, is incompatible with his recognition elsewhere of the self-reflective, emancipatory character of psychoanalytic therapy. He notes how Freud tried to define psychological concepts in accordance with this energy-model. Thus, for example, pain and pleasure are conceptualized in terms of the build-up and discharge of energy, and the overall aim of the psychic system is presented as reduction of the tension due to undischarged energy. Habermas claims that this approach led Freud, in his metapsychology, to characterize the function of the ego purely in terms of the economic management of energy. But this, he says, omits just what is revealed in analysis as the major capacity of the ego:

what does *not* appear among ego functions on the metapsychological level is the movement of reflection, which transforms one state into another - which transforms the pathological state of compulsion and self-deception into the state of superseded conflict and reconciliation with excommunicated language.⁷⁰

There is something that is correct and important in this criticism of Freud; but it does not require, as Habermas believes, that we reject Freud’s general commitment to the possibility of reductionism. That is, I suggest we should be critical of the particular way in which Freud specifies this reduction, and of how this affects his account of psychoanalytic theory, but not thereby of reductionism itself as defined in the previous section. The psychological reality of what Habermas calls ‘self-reflective processes’ should be recognized; and Freud’s energy-model at least conceals their distinctive character. But such recognition is compatible with reductionism though not, probably, with Freud’s version of it.

Before developing this response to Habermas, one qualification should be made to his view of Freud's rationale for the use of energy-concepts in characterizing psychological phenomena. It is arguable that Freud was not solely influenced by his early speculations about the energetics of the neurological system, but also by his early clinical experience, and perhaps by the widespread use of energy concepts in commonsense psychological statements. As Jahoda notes, there are:

a number of clinical observations which cry out for a language of psychic energy: 'the irrepressible nature of the neurotic symptoms (often voiced by the patient in such expressions as "There was something in me that was stronger than me"); the triggering-off of troubles of a neurotic kind following disturbances of sexual discharge . . .' etc. There is a parallel to this in the ineradicable notion of psychic energy in common-sense psychology. We speak of people needing or finding outlets for their energy or concentrating it on one area at the expense of others; when a child seems to have lost his zest we assume that this energy is invested in some secret preoccupation, etc. The need for a systematic approach to the circulation and distribution of psychic energy seems imperative.⁷¹

However, whilst agreeing with the general direction of these comments, I am sceptical about the implications of her last remark. For it suggests we must conceive of a system of energy-distribution in which, at any one time, there is a determinate total amount of energy quantitatively distributed in various ways (though whether this total is taken to be fixed or variable, and whether an accumulation-discharge model is assumed, is less clearly implied). I am not convinced that either the commonsense or clinical language of psychic energy should be thought to require these kinds of claims about an overall energy-system.⁷² None the less, I think it is a defect of Habermas's account of the therapeutic process that these energetic elements are markedly absent. This seems to be due to his almost exclusive concern with the linguistic-communicative features of therapy, and indeed of human activity in general. There is in Habermas's work a strong tendency towards the 'etherealization' of human practice, which leaves little room for its more somatic-experiential character rightly emphasized, for instance, in Reichian therapeutic techniques⁷³

Returning now to the specific nature of Freud's reductionism, I suggest that roughly speaking he worked the wrong way round. That is, he allowed the admittedly speculative neurological energy-model to influence strongly his conceptualization of psychological relationships, instead of first developing theories about these independently, and only then considering the possibility of their reduction to neurophysiology.⁷⁴ As Nagel has emphasized, the success of a scientific reduction requires that the reduced theory itself be fully articulated.⁷⁵ To try to construct it on the basis of its possible future reduction to an assumed reducing theory - especially when the latter is itself highly

speculative - is likely to be extremely misleading. For instance, even if the neurological energy-model were correct, it would not follow that the bridging-statements between its central principles and those of psychoanalytic theory must specify a simple correspondence between the concepts of energy in the two theories. This, whilst possibly attractive on the grounds of elegance or economy, may have little else to recommend it as a plausible theoretical strategy.

For this and other reasons (such as the difficulties Freud encountered concerning the quantitative basis for qualitatively distinct drives, and the implausibility of the tension-reduction goal assumption),⁷⁶ I support Habermas's view that psychoanalytic theory should be conceptualized independently of Freud's particular form of reductionism, or at least of some elements of this. And I agree that we should include amongst the functions of the ego the capacity for self-reflective dissolution of the unconscious determinants of compulsive activity. But such a view is compatible with the possibility of a neurophysiological reduction in which, amongst other things, we might hope to discover the bases in the central nervous system for this self-reflective capacity. Nor should we follow Habermas in characterizing the emancipatory process as a movement from causality to its abolition. We, as humans, have the ability to reflect upon our pasts, and to remove from their control over our lives certain unconscious influences. But the process through which this occurs is causally explicable; it involves a capacity that is biologically based and has emerged in the course of the evolutionary history of the species; and we should not assume that because we have this capacity our possible futures are limitless.

Notes and references

1 Jürgen Habermas, *Knowledge and Human Interests*, trans. J. J. Shapiro, Heinemann 1972, p.214.

2 Ibid., ch. 11 passim.

3 Ibid., pp.235-6. On the centrality of the theme of reconciliation in Hegel's philosophy, see Plant, Hegel.

4 Habermas, *Human Interests*, p.233.

5 Ibid., p.242.

6 Ibid., p.344, translator's note 31.

7 Sigmund Freud, *The Ego and the Id* (1923), p.25, in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, trans. and ed. J. Strachey, 24 volumes, Hogarth Press 1953-1974, vol. 19, pp. 1-66.

8 Ibid., p.25.

9 Ibid., p.56.

10 Ibid., 24.

11 Marie Jahoda, *Freud and the Dilemmas of Psychology*, Hogarth Press 1977, p.63.

12 See ch. 3, sec. 3, above.

13 See note 6 above. Freud's dictum was enunciated in his *New Introductory Lectures on Psychoanalysis* (1933), Penguin Books 1973, p.112.

14 Helpful guides to this problematic area of Freud's work are R. Fletcher, *Instinct in Man*, 2nd edn, Allen and Unwin 1968, ch. 5; and H. Nagera, ed., *Basic Psychoanalytic Concepts on the Theory of Instincts*, Allen and Unwin 1970.

15 See, e.g., S. Freud, *Introductory Lectures* (1917), Penguin Books 1973, lecture 1; Sigmund Freud, *The Interpretation of Dreams* (1900), Penguin Books 1976, pp.684-5; and the Editor's Introduction to this volume, pp.39-42.

16 Sigmund Freud, *Three Essays on the Theory of Sexuality* (1905), pp.82-3, in *On Sexuality*, Penguin Books 1977, pp. 33-169.

17 Freud, 'Instincts and their Vicissitudes' (1914), p. 123, in *Standard Edition* vol. 14, pp. 109-140; cf. the Editor's Note to this paper, pp.111-13.

18 Freud, 'The Unconscious' (1915), p.177, in *Standard Edition*, vol. 14, pp. 159-216.

19 Freud, *New Introductory Lectures*, pp.128-9; and cf. his comment: 'The concept of instinct is thus one of those lying on the frontier between the mental and the physical', *Three Essays on the Theory of Sexuality*, p.82.

20 On these changes see, e.g., Fletcher, *Instinct in Man*, pp.234-53; and Freud's own account in *New Introductory Lectures*, lecture 32.

21 On Freud's determinism, see *Introductory Lectures*, lectures 2 and 6; and his *New Introductory Lectures*, lecture 35.

22 Habermas, *Human Interests*, p.285.

23 Frankel, 'Habermas Talking', p.53.

24 Freud, 'Instincts and their Vicissitudes'.

25 See, e.g., Freud, 'The Unconscious'.

26 See especially *The Interpretation of Dreams*, pp. 745-69.

27 My account of the topographical model partly conflates several different stages in its development: see Freud, *New Introductory Lectures*, lecture 31, for a fuller account, including the relations between the topographical and structural models.

28 Freud, *The Interpretation of Dreams*, p.769.

29 *Ibid.*, ch. VI; *Introductory Lectures*, lecture 10; and *New Introductory Lectures*, lecture 29.

30 Freud, *The Interpretation of Dreams*, p.768. See also *Introductory Lectures*, p.339.

31 Freud, *The Interpretation of Dreams*, p.769.

32 *Ibid.*, p.763.

33 On p.3 above.

34 Freud, *The Ego and the Id*, p.56.

35 See Fletcher, *Instinct in Man*, p.198, where this parallel is noted.

36 See, e.g., *The Ego and the Id*, ch.V; *Introductory Lectures*, lecture 28; and *New Introductory Lectures*, lecture 34.

37 See especially Sigmund Freud, *Civilization and its Discontents* (1930), *Standard Edition*, vol. 21, pp. 49-145. As is often noted, e.g. in Paul Robinson, *The Sexual Radicals*, Paladin 1972, pp.19-63, Freud's introduction of the death instinct into his theory led to a marked change of attitude towards the social consequences of sexual repression from that expressed in, e.g. his earlier "'Civilized" Sexual Morality and Modern Nervous Illness' (1908), *Standard Edition*, vol. 9, pp.177-204. Rejection of the death instinct was a major reason for Reich's disenchantment with Freud: see, e.g. Wilhelm Reich, *The Function of the Orgasm* (1927), Pocket Books 1975, ch. 5, and *The Sexual Revolution*, 4th edn, Vision Press 1972, Part I, ch. 1.

38 Habermas, *Human Interests*, p.271.

39 *Ibid.*, p.272.

40 This issue is returned to briefly in ch. 6, sec. 2, below, in discussion Habermas's conception of an 'ideal speech situation'.

41 My classification here roughly follows that given in the Introduction to Bernard Berofsky, ed., *Freewill and Determinism*, Harper and Row 1966. See also Ted Honderich, *Punishment: The Supposed Justifications*, Hutchinson 1969, ch. 4.

42 On this last issue, see C. V. Borst, ed., *The Mind-Brain Identity Theory*, Macmillan 1970, pp.13-33. For a discussion of both this, and the logical status of purposive explanation, see Jerome Shaffer, *The Philosophy of Mind*, Prentice-Hall 1968, pp. 39-50 and 77-106.

43 Both points are defended in Keat and Urry, *Social Theory as Science*, pp.151-9.

44 Cf. Brian Fay, 'How People Change Themselves' (in Terence Ball, ed., *Political Theory and Praxis*, Minnesota University Press 1977), who adopts a fairly similar strategy in this respect.

45 Habermas, *Human Interests*, p.271.

46 *Ibid.*, pp. 216-17.

47 See Habermas's account of Dilthey, in *Human Interests*, ch.7.

48 See note 37 above. Curiously Marcuse, in *Eros and Civilization* (2nd edn, Beacon Press 1966), produced a critical account of the social function of sexual repression quite similar to Reich's, but accepted Freud's death instinct. I am doubtful about the consistency of this position.

49 See, e.g., Margaret Boden, *Purposive Explanation in Psychology*, Harvard University Press 1972, ch.V; and Mary Midgley, *Beast and Man*, Harvester Press 1978, ch.3. See also the references in note 59 below.

50 For a discussion of this question, see Boden, *Purposive Explanation*, ch. VI.

51 Ernest Nagel, *The Structure of Science*, Routledge and Kegan Paul 1961, ch. 11. See also Boden, *Purposive Explanation*, pp.52-9; what I go on to call 'explanatory' reduction corresponds to Boden's 'empirical' reduction; and her concept of 'strict' reduction includes what I call 'ontological' and 'conceptual' reduction.

52 See ch. 3, sec. 1, above.

53 For an account of materialist and other 'solutions', see Shaffer, *The Philosophy of Mind*, chs. 3 and 4.

54 Nagel, *The Structure of Science*, pp. 364-6.

55 *Ibid.*, p.152.

56 For a criticism of Nagel's view, in *The Structure of Science*, ch. 6, see Grover Maxwell, 'On the Ontological Status of Theoretical Entities', in H. Feigl and G. Maxwell eds, *Minnesota Studies in the Philosophy of Science*,

vol. 3, *University of Minnesota Press* 1962, pp.3-27.

57 See, e.g. Shaffer, *The Philosophy of Mind*, pp.68-71.

58 See Steven Rose, *The Conscious Brain*, rev. edn, Penguin Books 1976, pp. 61-94, for a standard account of the neuron.

59 See Rose, *The Conscious Brain*, ch.8; and S. A. Barnett, 'Instinct' and 'Intelligence', Penguin Books 1970, chs. 9 and 10, for useful discussions of these issues; also note 49 above.

60 S. Freud and J. Breuer, *Studies in Hysteria* (1895), Penguin Books 1974, p.252.

61 See, e.g. Reich, *The Function of the Orgasm*, ch. 7. For a standard account of the autonomic and voluntary nervous systems, see E. R. Hilgard et al., *Introduction to Psychology*, 6th edn, Harcourt Brace Jovanovich 1975, ch. 2.

62 See Rose, *The Conscious Brain*, ch.4.

63 See Richard Gregory, *Eye and Brain*, 3rd edn, Weidenfeld and Nicholson 1977, especially chs. 7, 9, 11 and 12.

64 See Rose, *The Conscious Brain*, pp. 280-4.

65 Boden, *Purposive Explanation*, pp.154-5.

66 Ibid., p.155.

67 Habermas, *Human Interests*, p.247.

68 Freud, *An Outline of Psychoanalysis* (1938), p. 182, in *Standard Edition*, vol. 23, pp. 139-207.

69 Freud, *Introductory Lectures*, p.487.

70 Habermas, *Human Interests*, p.245.

71 Jahoda, *Freud*, p.101. The quotation early in this passage is from J. Laplanche and J. B. Pontalis, *The Language of Psychoanalysis*, Hogarth Press 1973, p.247.

72 See, e.g. the critical discussion of energy-discharge models in ethology in A. Manning, *An Introduction to Animal Behaviour*, 2nd edn, Edward Arnold 1972, ch. 4.

73 See ch. 6, sec. 1, below.

74 For Freud's early formulation of this energy-model, see his *Project for a Scientific Psychology* (1895), *Standard Edition*, vol. 1, pp. 283-346.

75 Nagel, *The Structure of Science*, p.345.

76 On these issues, see Fletcher, *Instinct in Man*, pp.174-83 and 239-53.