Note to preceed the re-printed Introduction to Philosophical Essays on Freud

The Introduction to Philosophical Essays on Freud was written to deadline over a decade ago, and the flaws in my statements of both philosophical ideas and clinical descriptions are more glaring with time. Still the account seems worth going on with, and the essay remains the only short philosophical introduction to these topics. I have revised and extended a number of ideas in subsequent articles, and also written a separate discussion of the work of Melanie Klein, which is treated mainly in footnotes. These developments flow fairly naturally from the text here, and are noted in the comments on Johnston which follow.

Irrationality, Interpretation and Division: Comments on Mark Johnston's Essay

A main theme of the Introduction to Philosophical Essays on Freud was that a significant part of Freud's thinking could be understood in terms of the notion of wishfulfilment, or wishfulfilling phantasy; and that this should be regarded not as intentional action, but rather as a form of wishful thinking or imagining, in which a wish or desire causes an imaginative representation of its fulfilment, which is experience- or belief-like. Such a causal sequence has a pattern, which it will be useful to set out more explicitly, so that it can be compared with others. Letting the agent be A, and abbreviating the notion of belief- or experience-like representation by 'b-rep', we can write the pattern as

(1) A's desire that P -[causes]-> A's b-rep that P

Simple instances of this pattern were the case in which (as we can put it) Freud's desire that he drink causes his dream that he is drinking (xx), and also the symptom of the unembarrassed girl (xi), caused, there seems reason to say, by an underlying desire which was sexual. In these cases the imaginative representation has a content which is easy to grasp. Hence the content of the underlying causally active desire (or wish) can be read from its effect in accord with (1), which serves as a sort of template for interpretation of this kind.

In more complex instances -- such as the dream of not giving a supper party with smoked salmon (xx), or again the obsessional symptom of showing the maid the spot (xxi) -- the representational content is less manifest. Where this is so the content must be brought more fully to light, by way of the free associations of the patient or dreamer, before the pattern can be applied. Still the pattern of even very complex examples remains simple in form, and has been traced in empirical material in a vast range of instances and cases, often in remarkable detail. So (1) can be seen as having a pervasive role in psychology, despite both its internal simplicity and the interpretive complexity through which some of its instances are disclosed.

Although Mark Johnston's 'Self-Deception and the Nature of the Mind' was not written with these claims in mind, it seems at least in part to accord with them. In particular, Johnston independently emphasizes the role of wishful thought, and also argues that this should be understood as involving causation which is non-rational and non-intentional, and which holds between desire and something like belief. He describes the causal role of desire in such cases in terms of 'mental tropism', and speaks of the result as 'quasi-belief', which seems close to the idea of belief-like representation above; and he also takes the tropistic causation of quasi-belief to be very common. Johnston also indicates in passing that he too applies, or would apply, this account, not only to self-deception, but also to a number of phenomena described in psychoanalysis: division, denial, repression, removal of affect, and wishful perception and memory. Thus we seem to agree on these main points, and also to share a common perspective on the philosophical exposition of psychoanalytic theory.

Johnston also sets out a number of other lines of argument. He holds, for example, that the commonness of wishful tropism contradicts Davidson's intepretive approach to the mind; and he argues, as against both Davidson and Pears, that a tropistic account of self-deception avoids difficulties inherent in their homuncular ones. Despite the interest and force of the arguments which Johnston puts foreward, these conclusions seem to me premature, and to require serious qualification. In what follows I will concentrate on these points of disagreement, and especially as they bear on the philosophical understanding of psychoanalysis, which I shall discuss in as much detail as space permits. The reader should bear in mind, however, that this was not Johnston's main topic, and hence that emphases and additions in my discussion are not meant to indicate shortcomings in his.

II

Let us first take Johnston's critique of Davidson, whose approach to commonsense psychology I was attempting to extend to psychoanalysis. As Johnston observes (p 80), Davidson says in 'The Paradoxes of Irrationality' that

> the only clear pattern of explanation that applies to the mental...demands that [a] cause be more than a candidate for being a reason; it must be a reason [for what it causes]

Now plainly Johnston and I disagree with Davidson on this point, since we both take the causal pattern of wishfulfilment above to be clear, explanatory and widely applicable to the mental. Since Davidson is also well aware of this pattern, and indeed calls it 'a model for the simplest form of irrationality', I have thought the statement above a slip, with little bearing on his main views. Johnston, however, regards it as more consequential. He writes that

> The existence and ubiquity of mental tropisms whose relata to not stand in any rational relation falsifies a view of the mental which is gaining currency. This interpretive view of mental states and events has it that there is nothing more to being in a mental state or undergoing a mental change than being apt to have that state or change attributed to one within an adequate interpretive theory, i.e. a theory that take's one's behaviour (including speech behaviour) as evidence and develops under the holistic constraint of constructing much of that behaviour as intentional action caused by rationalizing beliefs and desires that it is reasonable to suppose the subject has, given his environment and basic drives (66).

Davidson's approach can rightly be called interpretive, but this description does not do it justice. Thus it is doubtful that according to Davidson there is 'nothing more to being in a mental state or undergoing a mental change' than being apt to have that state or change interpretively ascribed. For Davidson holds that mental events are identical with physical events, that mental properties (predicates) supervene on physical ones, and that a person's psychological dispositions and abilities, which include desires, beliefs and the capacity to speak and act intentionally, are realized or 'constituted' by 'physical state[s], largely centered in the brain'. In consequence Davidson notes that in identifying a physical event with an action, say, we must 'be sure that the causal history of the physical event includes events or states identical with the desires and cognitive states that yield a psychological explanation of the action.' Thus Davidson explicitly constrains the role of interpretation in his account of the mental by a series of notions -- identity, supervenience, and constitution or realization -- which relate the mental to the physical, and so as to ensure that all causal relations of the mental have an appropriate physical realization 'centered in the brain'. The role of these physical constraints has not been fully articulated, but their existence contradicts a literal reading of Johnston's 'nothing more' above.

As is well known, Davidson also argues that no strict law -- no law which is sharp and exceptionless, and which contains no ineliminable caveats, ceteris paribus clauses, or the like -- holds between a particular type of desire and any type of constituting state or mechanism in the brain. Thus, e.g. the type desire to have coffee with milk cannot be strictly connected with any neural mechanism which can be precisely specified in physical terms. This can be seen as a direct consequnce of his emphasis on interpretation. The idea would be that if we hold that the ascription of a desire is ultimately answerable to the interpretive explanation of behaviour, then we cannot at the same time hold that the desire is related by a strict law to some well-defined physical mechanism. For the presence or absence of such a mechanism would be a clear physiological fact, and this, by the strict law, would fix the presence or absence of the desire in all nomologically possible circumstances. So to hold that there is such a mechanism-specifying law would be (implicitly) to take the ascription of the desire as answerable to a specific mechanism rather than to the explanation of behaviour.

In envisaging the existence of psychophysical laws we tend to assume that everything will come out in perfect harmony: that there will be no conflict between ascribing the desire on the basis of its supposedly lawfully constituting physical state or mechanism, and ascribing it on the basis of interpretation; nor any uncertainty, given the mechanism, as to whether the desire is actually there. This, however, misses the point of the above argument, which is that the existence of strict law should actually preclude the empirical possibility of this kind of disharmony. In holding that ascriptions are ultimately answerable to considerations of interpretive psychological explanation we show that we do not assume that this possibility is foreclosed, and hold that the final verdict lies with interpretation. This, however, is the position for which Davidson argues.

Another argument may bring out the nature and plausibility of this position, and will also serve to introduce some further matters relevant to the discussion. Let us reflect on our practice of describing motives, which we schematize by speaking of the desire that P, the belief that Q, and so forth. In this 'P' and 'Q' stand for, or can be replaced by, sentences of natural language. We understand these sentences, in turn, as true in the worldly conditions or situations which they specify, and hence in accord with a semantic pattern which we can indicate by

(2) 'P' is true just if P

This schematic pattern is supposed to cover our systematic understanding of the truth-conditions of indefinitely many sentences of our language, and hence to describe a vast amount of information relating language and the world -- which, of course, it does only very roughly indeed. Equally schematically, the conditions in which we take the sentence 'P' to be true are also those in which we take the desire that P to be satisfied, the belief that P to be verified, the hope or fear that P to be realized, and so on. Thus to take our example: when we say that Freud desired that he (Freud) drink, we use the sentence 'he (Freud) drink[s]' to describe Freud's desire; in accord with pattern (2), we take 'he (Freud) drinks' to be true just if Freud drinks; this, therefore, we also take as the situation in which Freud's desire would be satisfied. So in accord with (1), this is the situation which Freud b-reps as obtaining, in dreaming that he (Freud) drinks. The artificiality in these phrasings reflects something of the roughness in our specification of the relevant patterns; but the existence of such patterns seems clear.

Our commonsense practice is thus to recycle our worldly sentences, to describe the mind in its engagement with the world; and we do this in such a way as to enable us to understand this engagement in accord with our understanding of the sentences themselves. This mode of description is at once semantic and causal: for in employing it we use the semantic relations of our motive-describing sentences to map causal relations between motive and motive, and motives and the world. We take it, for example, that desires serve to bring about (cause) the semantically specified conditions in which they are satisfied. This shows in a further pattern, which we can write as follows:

(3) A's desire that P -[causes]-> P

This is a pattern we find in successful rational action, such as that in which Freud desire is that he drink, and this brings it about (causes) that he drinks. This form, incidentally, is common to other kinds of teleological explanation, which postulate representations of goals which operate within the systems of which they are part to bring about (cause) those goals. What is unique about explanation by desires is not this basic teleological form, but rather that the system in which goals (desires) are represented has the expressive and computational power of human language (cf the pervasive role of (2) above). Also, of course, this pattern can still be applied when the connection between desire and situation is mediated by further desires and beliefs, as discussed below.

Similarly, we hold that beliefs serve to register (be caused by) the situations which verify them. This gives the pattern

(4) P -[causes]-> A's belief that P

This pattern is also widespread: we take it, for example, that when Freud drinks, this brings it about that he believes that he drinks; and something like this is characteristic of intentional actions generally. Again, this pattern is rational, since it is that of belief which is both true and justified by the presence of a kind of causal relation which makes for knowledge. Like (3), this pattern extends to cases in which the connection between belief and situation is mediated, e.g. by further belief or theory. Indeed it might be said that the point of theory is to make our beliefs sensitive to the world as in (4), so that our desires can work in it as in (3).

We also use deductive semantic relations between sentence and sentence to map causal relations between motive and motive. In general, we take it that beliefs cause beliefs in accord with logical patterns (not specified here). Also we apply such patterns to desires, for example in holding that an agent who desires that Q and believes that if P then Q thereby has reason to desire that P. This is the causal pattern of practical reason:

(5) A's desire that Q and belief that if P then Q -[causes]-> A's desire that P

Here the pattern of motive-specifying sentences, read from right to left, is that of modus ponens. This shows that if the belief in the pattern is true, then satisfying the derived desire must also satisfy the initial one, so that an agent who forms or modifies motives in accord with the pattern is thus far rational.

(3) and (4) relate motives described by sentences to the worldly situations

in which those sentences are true, and so in effect incorporate (2); so the patterns with which we are dealing are at once semantic, rational and causal. They illustrate how our commonsense system of understanding persons co-ordinates norms of language, as partly specified in (2), with norms for the working of motive, as specified in (3) - (5); so that the understanding of language, and that of motive in rational action, form a hermeneutic and causal unity. Hence, arguably, the basic form of application of this system is in the interpretation of persons as rational, i.e. as thinking, acting, and speaking in accord with such norms, and thus in accord with schema like (2) - (5).

Such interpretation can be represented as a process in which an interpreter explains sequences of an interpretee's bodily movements as intentional actions, motivated by desires and beliefs with appropriate environmental conditions of truth, satisfaction, and the rest, where this includes interpreting the making of strings of sounds as the utterance of sentences with particular environmental conditions of truth. In this the interpreter in effect maps sentences of his or her own language on to both the behaviour of the interpretee and the environment shared by them both, and thereby systematically links the interpretee's behaviour with that common environment.

In the simplest case, in which the interpretee happens to use the same words and sentences as the interpreter, and in the same way, interpretation can at least partly be understood in terms of the repeated application of patterns like (2) - (5); for in the idealized situation in which the interpretee judges accurately and so acts successfully, the interpreter can always use a single sentence, or closely related sentences, to characterize both the interpretee's belief and its object, both his desire and its satisfaction, and both his sentence and its meaning. Such overlapping characterization, in turn, registers that the causal relations which hold among the interpretee's motives and the environment generally are as they should be, and this is the simplest case of the kind of pattern of coherence which marks successful psychological explanation. The interpreter's own use of language, including that in interpretation, is also to be understood as construable by interpretation, and hence answerable to the norms imposed in the course of it, in the same way as the interpretee's. So an interpretive view offers us an account of the content and causal role of both motives and sentences, as fixed in harmony through the applicability of interpretive teleological (and causal) explanation of behaviour.

It seems, therefore, that interpretive patterns such as those indicated in (2) - (5) have an epistemic status which is worth noting. We interpret behaviour in accord with them naturally, and hence spontaneously, rapidly, and continually. In this sense we use such patterns more frequently, and rely on them more deeply, than any generalizations of science. (But of course we have no need to realize that this is so.) We apparently learn such patterns together with language, so that their use is in a sense a priori. Also, however, we find them actually instantiated, and hence supported in a way which is both empirical and a posteriori, in instances of successful interpretive understanding too dense and numerous to register.

Patterns of this kind are also predictive. For example when an interpreter takes it that an interpretee is acting, or is going to act, on a certain desire, the interpreter's description of the desire by a sentence 'P' constitutes a prediction in accord with (3). The interpreter's description, that is, can be regarded as an hypothesis, which is framed and tested by successive uses of the same sentence: the first use describes the motive, and the second the action or situation which this motive should bring about. (Roughly, the hypothesis is confirmed if the sentence used to describe the desire also serves to describe the action or situation caused by the desire, and disconfirmed if not; hence, as stressed in the introduction, the hypothesis tends to be confirmed by sentential coherence, and disconfirmed by its absence.) Something analogous also holds in instances of (1), (4) and (5); so that uses of these patterns constitute a system which is subject to a variety of predictive tests, each framed by the use of a single sentence and confirmed or disconfirmed by further uses of that sentence, and hence by instances of descriptive (interpretive) coherence, or lack of it. More generally, success in interpretation enables us to achieve and predict further success of the same kind. So past use of such patterns gives us good reason to count on finding them instantiated in the future, and on their continuing to locate the same sorts of motives that they have consistently specified so far.

Part of the argument of the introduction -- that involving the 'strongly predictive guiding principle[s] of interpretation', of action and wishfulfilment sketched at xxvi ff -- can be represented as claiming that (1) has acquired an empirical status akin to that of (3), through the psychoanalytic interpretation of episodes in behaviour which had not previously been observed or understood; and that in this use (1) and (3) tend in fact to home in on the same recently discovered but basic motives. Uses of (1) and (3) thus interact in psychoanalytic practice to support one another in an extension of commonsense psychology which is potentially sound, cumulative, and radical. Sound, because the extending interpretations cohere both with the basic patterns, and also with one another, in locating very many supporting instances for the relevant values of P; cumulative, because each discovery of the operation of new motives naturally facilitates the discovery of others; and radical, because the extension offers significantly deeper, fuller, and more coherent explanations of actions and wishfulfilments generally, and by reference to motives which, in the main, had not previously been contemplated. This view still seems to me to be mainly correct, but will receive some revision in what follows.

We have seen that (2) - (5) can be taken to describe patterns of psychological dispositions -- to link sentence with situation, situation with motive, and motive with motive -- which accord with causal and semantic norms. Hence interpretation in accord with such patterns represents the mind of the interpretee as rational, and as a semantic engine, whose inputs, working, and output are registered in terms of belief, reason, and the satisfaction of desire. The mechanism which realizes such dispositions, and hence the engine which we thus indirectly describe, is the nervous system, and in particular the brain. Patricia Churchland takes the aim of research in computational neurophysiology to be that of mapping the 'phase-spaces' -- the 'as the world presents itself' space, and the 'as my body should be' space -- which the working brain relates. This, it seems, is also the task which commonsense psychology already partly performs, via the use of natural language, in describing persons in terms of patterns like (2) - (5).

Such causal-semantic description of motives is like our commonsense description of a photograph, which doesn't describe the picture chemically, say, but rather in terms of the objects or persons in the environment which are represented in it, and which played a certain causal role in its production. We assume that the look of a photograph supervenes on its intrinsic physical state, and also that this state can be explained causally, by reference to the objects or situations specified in an environmental description and the physical processes involved in photography. Environmental and intrinsic descriptions of photographs are in a clear sense descriptions of the same things (the same representations). They are both useful, and further scientific inquiry can specify them further and relate them in greater detail; and since they are not competitors no sensible person who knew their uses could want to eliminate either. Also we know that the same environment can photograph in different ways, and different environments in the same way, so we take it that such descriptions will not be strictly related type by type.

This situation seems entirely unproblematic in the case of representations like photographs. So if we take the desire or belief that P to involve a neural representation of the situation P, the parallel claim should be equally acceptable. This, however, is the claim that there is no strict correlation between psychological descriptions (that P descriptions) and intrinsic physical descriptions of representational mental states. We can be sure that an environmental description of a representation will be systematically connected with intrinsic descriptions of the same representation, when we can frame them. But describing representations in terms of a complex causal role vis-a-vis the environment is very different from describing them intrinsically. Knowing the working of these two forms of description -- and in particular knowing that description via causal connection with the environment invariably introduces a degree of slack -- we can see that their correlation should not be strict.

Both this argument and that derived from Davidson above turn on the notions of interpretation and strictness. States which are described by the interpretive specification of environmental conditions of truth, satisfaction, and the like will not be strictly related to specific internal mechanisms. As the analogy with photographs suggests, however, the slack need not be great. Thus it is consistent with the letter of such arguments that there there should be something like a language of thought, with a specific 'syntactic' neural mechanism for each desire or belief -- so long as this syntax was, say, ineliminably rough, or susceptible to ambiguity, local variation, or the like. Still acknowledgment of the role of interpretation suggests something more radical. As a connectionist might put the point, it seems that we should take desires and beliefs as realized by neural networks which may vary from person to person, or, even in the same person, from time to time. This also allows that the network realizing one desire or ability can realize others, so that there would be no strict pairing of distinct parts of the network, or of distinct neural structures, with distinct desires or beliefs. This is not part of Davidson's argument, but fits with his emphasis on the interconnection of beliefs and other attitudes with content. For on this picture -- a version of which was urged at xv, footnote 9 -- the agent's system of motives, and their neural realization, would match not item by item, but rather only net by net.

Johnston goes on to claim that

> ...On this conception, when we attribute a mental state to another, we are not locating within him an instance of a mental natural kind or property that as such enters into characteristic causal relations in accord with nonaccidental psychological or psychophysical regularities. On the view in question there are no natural mental properties and so no lawlike psychological or psychophysical regularities. Instead, attributions of mental states and changes have point only within a whole pattern of reason-explanations, i.e. explanations that exhibit the subject as a rational agent pursuing what is reasonable from his point of view. Fitting into a pattern of reason-explanations that serve to interpret their subject is thus a constitutive condition of something's being a mental attribution. More, there can be no other content to the idea that something is a mental attribution. In this sense, rationality is constitutive and exhaustive of the mental (66).

Thus, as Johnston spells out his refutation of Davidson's view:

> ...wishful and self-deceptive thought seems to involve a characteristic and explanatory causal connection between the desire that p and the belief that p, but an explanatory connection which is not a rational connection. The anxious desire that p is not a reason to believe that p. Because the interpretive view counts rationality as both constitutive and exhaustive of the mental, it has trouble finding a place for the very possibility of a mental state, anxious desire, which characteristically has irrational mental consequences (80).

Again, however, Johnston's remarks do not characterize Davidson's view correctly. Since Davidson's argument is directed only against strict laws, it is consistent with a range of claims about realization. And even if we take the radical alternative sketched above, it remains false to say that we are not, in attributing a mental state to another, locating within that other something which 'as such enters into characteristic causal relations in accord with non-accidental psychological or psychophysical regularities.' For on any such view mental states like desires and beliefs still enter 'as such' into characteristic causal relations, and in accord with non-accidental psychological regularities. Such states involve dispositions and Davidson suggests that 'the laws implicit in reason explanations are simply the generalizations implied by the attributions of dispositions' of this kind. The attribution of dispositions (or causal powers, capacities, etc.) is clearly that of non-accidental generalizations. To ascribe a desire is to attribute, among other things, a disposition to produce a situation in which the desire is satisfied. So here the relevant non-accidental psychological regularity is that specified in (3) above. (And of course we also locate the mechanism which realizes the disposition 'within' the person to whom we ascribe the desire.) This, however, seems a companion pattern to the version of (1) which Johnston also describes. Davidson's slip apart, these patterns seem too alike in role and content to be distinguished further to his disadvantage, despite the fact that (3) is rational while (1) is not.

As noted, Davidson has stressed that such generalization over desire and action as we find in (3) is not strict. The predictive use of (3) considered above, for example, involves a claim that an agent is acting, or will act, on a specified desire. But the mere fact that someone has a desire -- even a strong desire -- does not itself render action on that desire interestingly probable. Countervailing desires may usually, or invariably, be stronger, and for many desires we may know this in advance. Davidson gives the example of 'the ratio of actual adultries to the adultries which the Bible says are committed in the heart'; and despite their strength, persistence, and pervasive influence, the ratio for intentional action on Oedipal desires is more infinitesimal yet. In these cases we both accept that the desires involve dispositions specified in terms of action and also hold that such action is very unlikely; that is, we take the desires to show themselves via (1) rather than (3). Also, as Davidson emphasizes, we generally have no way of specifying in advance of action which desires -- which values for P in (3) -- will be strongest at a given time, or which will get acted in accord with. This point also holds for (1); and the fact that both generalizations have the same antecedant, desire, suggests that their lack of strictness is comparable.

This lack is no sign that we cannot interpret actions and wishfulfilments in accord with these patterns accurately and efficiently. Our natural interpretive abilities would seem to have arisen because they enable us to extract information from the behaviour of others (cf xix), and hence to have evolved together with the forms of behaviour which they enable us to understand. The patterns are thus made for the sequences of behavior on which they operate, and vice-versa. The cognitive task of extracting information from such behaviour is not solely that of saying ahead of time what the sequences will be; the point is not just to predict the information-bearing specifics, but rather to use them, and hence the information they carry, for further purposes (which may include prediction of other things). Thus the lack of predictive strictness in the patterns is no fault, but rather a mark of their fitness to the task we perform in accord with them.

Moreover, the pattern of the working of desire in successful action should not be seen as distinct from that shown in wishfulfilment. Both, rather, are implicit together in rational action itself. To bring this out let us distinguish between the satisfaction and the pacification of a desire, as follows. A desire is satisfied just if its conditions of satisfaction obtain, and in particular if it operates to secure these in successful action; and a desire is pacified if it is caused to cease to operate, or to alter in its operation, in certain normal ways. In terms of this distinction, we can say that in the everyday explanation of action, we assume that satisfaction characteristically causes pacification, and at least partly by way of belief. Thus Freud desires that he drink, and as a result he drinks, so that his desire is satisfied; and as a further result he believes that he has drunk, and this belief, perhaps together with the drink itself, pacifies his desire to drink, so that it ceases to govern his actions. In wishfulfilment, by contrast, we take desire to cause satisfaction-like experience and quasi-belief directly, and so to yield pacification without satisfaction. Thus Freud desires to drink, and as a result dreams that he is drinking, and this, as it seems, pacifies his desire, at least temporarily.

The phenomenon of pacification as distinct from satisfaction plays a salient role in interpretive practice. One of our principal ways of verifying hypotheses about the desires upon which we presume people to be acting is through observing that they cease to act as on a particular desire precisely when that desire should be pacified -- that is, when the relevant conditions of satisfaction obtain, and the agents become aware of this. We thus implicitly interpret action by reference to the content of pacifying belief, just as we explicitly interpret wishfulfilment by reference to the content of pacifying quasi-belief; and interpretation in both cases consists in taking the believed content as derived from a desire which the action or wishfulfilment serves to pacify. In light of the phenomenon of pacification, we must note that action and representation are fundamentally interwoven; action and phantasy are both aimed at the production of pacifying representation, and interpreted by reference to the content of this.

Taking pacification explicitly into account, we can say that our understanding of the causal pattern in even the simplest successful action involves a fuller pattern than that which appears in (3), which we can write as:

(3)* A's desire that P -[causes]-> P -[causes]-> A's belief that P -[causes]-> A's desire that P is pacified

As before, this can serve as a predictive pattern, in which we frame a hypothesis by the use of the sentence which describes the desire, and test this hypothesis by further uses of that same sentence, to characterize not only the action or situation which the desire brings about, but also a belief which the agent forms as a result of this, and the consequent inner process, in which the belief causes the initial desire to cease to operate. (This cycle, again, is found in other forms of teleological explanation, in which a representation causes a system to attain a goal and this in turn causes a further representation which operates to curtail or alter the activity of the first.) In the case of human motive the pattern includes within that of veridical belief, specified in (4) above; and in this it particularly contrasts with wishfulfilment. (1) can likewise be filled out to

(1)* A's desire that P -[causes]-> A's b-rep that P -[causes]-> A's desire that P is pacified.

This makes it clear that we can regard wishfulfilment as a kind of short-circuiting of a path from desire through reality to belief and pacification which is at every step rational, and which we find in successful action. For we can see (1)* as obtained from (3)* precisely by the omission from the latter of the causal role of the situation that P, which is real and satisfying, and causation by which distinguishes veridical belief from belief-like representation. This emerges again if we note that (3)* is cast in terms of belief, and (1)* in terms of belief-like representation. Since belief can be treated as the limiting case of belief-like representation, we can rewrite (3)* in a more general form as

(3)** A's desire that P -[causes]-> P -[causes]-> A's b-rep that P -[causes]-> A's desire that P is pacified

Again, omitting the causal role of the real situation that P yields (1)*, the pattern of pacifying wishfulfilment.

It thus appears that contrary to Johnston's statement above an interpretive view of the mind -- even one which places particular emphasis on the 'pattern of reason-explanations' -- need have no difficulty in finding a place for the pattern relating desire to quasi-belief. For both this latter pattern and that relating desire to satisfaction now appear as partial sub-patterns in the overarching and representation-mediated connection between desire and pacification which is characteristic of rational action itself.

In this perspective the interpretation of rational action and wishfulfilment are naturally interrelated. In commonsense psychology we interpret actions in accord with the basic generalization that the role of a desire that P is to produce a situation that P, which in turn should produce a belief that P which (perhaps together with the situation) pacifies the desire. In understanding persons we both tacitly use this generalization, and also sustain it inductively, as noted above. Since this generalization already includes the idea that representation (belief) that P serves to pacify the desire that P, we also take it as an intelligible, and indeed common, phenomenon that a desire that P should play a role in causing a belief-like representation that P, which tends to pacify the desire. This is, indeed, another generalization which we already both use and sustain, in understanding many forms of pacifying representation with which we are familiar. These include a variety of kinds of children's play, and adult representations such as those of literature, art, cinema, and such related achievements as advertising and pornography. We know, of course, that pacification consequent on real satisfaction and veridical belief is, among other things, more thorough and lasting than that obtained through representation or phantasy. But we also know that desire far outruns the possibilities of satisfying action, so that attempts at pacification by representation alone are common.

Thus I think we already understand, say, that a child may represent itself as a hero in play, or that we repeatedly represent certain situations in fiction, film, etc., because these situations seem desirable, and their representation therefore provides opportunity for the pacification of desire, via one form or another of quasi-belief (cf the notions of make-believe, suspension of disbelief, and so forth). We are aware, e.g., that someone playing a video game in which he mutilates a variety of enemies may not only be satisfying a desire to play a game, but also pacifying other desires which the game represents as fulfilled; and that the arousal and pacification of these desires may be a source of the excitement of the game. Understanding of this commonsense kind is continuous with the psychoanalytic interpretation of a dream or symptom. Thus compare Freud's interpretation of the Rat Man's recurrent symptom of imagining his father being punished by the rats, and feeling anxiety and depression as a result. In accord with (1)* we can see this as expressing a wish that his father be punished, repeatedly represented as satisfied, and therefore repeatedly temporarily pacified, in the virtual reality of his own phantasy, whose boundaries he could not keep distinct from everyday belief. The patterns with which we began can thus be regarded as implicit and interwoven in pre-theoretical commonsense understanding of action and representation, and hence also as capable of interacting to extend commonsense psychology, in something like the way indicated above.

III

Let us now turn to the division of the mind, as this appears in Davidson and Freud. Here, I think, we encounter what can be seen as two distinct tendencies of thought. Davidson's divisions in the self are 'overlapping territories' in the field of an agent's motives: they are 'constellations of beliefs, purposes, affects' which co-operate rationally with one another in producing intentions; and they can conflict with other such constellations, or motives in these, and in this act 'in the modality of non-rational causality'. Strictly speaking such constellations do not have motives; rather they are (groups or families of) motives, many of which have a role in more than one constellation, and all of which are had by one and the same agent. So they are not really distinct agents, but at best analogous to these. Davidson stresses explicitly that 'The analogy does not have to be carried so far as to demand that we speak of parts of the mind as independent agents...the idea of quasi-autonomous division is not one that demands a little agent in the division.' His idea thus seems to be to stop short of the postulation of homunculi, and make do with cohesive groups of motives instead. And it is difficult to see how the explanation of irrationality might proceed without reference to such groups in any case (how, for example, might the wish not to know cause one to avoid relevant evidence, except via the web of belief?)

Freud's ego, super-ego, and id, by contrast, can certainly be regarded as distinct and autonomous 'agencies'. Insofar as this is so, however, they seem not best thought of as agents which have desires, beliefs, and practical reason, but rather as functional systems, which we describe in a teleological way, that is, in terms of the goals which we take them to operate to secure, and the information they use in doing so. This seems the way to interpret, e.g. Freud's description of the ego as 'a special organization... which acts as an intermediary between the id and the external world', and which also 'makes far-reaching changes in its organization' in the state of sleep; or again his description of the super-ego as 'a special agency [in the ego] in which...parental influence is prolonged' (XXIII, 145,146). Teleological description of this kind is closely related to that in terms of beliefs and desires; but the two have differences which are relevant to the present discussion.

As noted above, when we describe people in terms of desires and beliefs, we can also be regarded as indirectly be describing a neural system (the human brain) in a teleological way, in terms of the environmental goals of the system and information upon which it operates. In this, however, we represent the goals and information in terms of human language, and thereby imply that the system (person) we are describing represents goals and information in a comparably subtle and powerful way. We take it that a person to whom we ascribe a desire for Scotch Whiskey, for example, has the concept of Scotch Whiskey, and therefore has many beliefs about Scotch Whiskey and can readily compute many more (cf xii); and we make constant tacit use of this in interpretation, for example in our applications of logical principles such as (5). Where we give teleological explanations of the behaviour of animals and artifacts, however, we relax this implication. Thus we take it that a rat whose goal is to get Scotch Whiskey -- say by pressing a bar -- has some representation of this outcome, otherwise we would not acribe this goal. But we do not assume that the rat has our representation of Scotch Whiskey, and so we do not regard the ascription of the goal as having the same consequences as in the case of a person.

The point is the same in the case of the ego, super-ego, and id. We may describe such systems as if they had motives, in describing their goals, and information on which we take them to operate; but we do not take these descriptions to have the same consequences as in the case of the desires and beliefs of persons. And since the motives of persons are the basic paradigms of desire and belief, we might do better to follow Wittgenstein and Davidson, and say that in such non-paradigmatic uses the animals (or artifacts, or systems of neurons) do not have desires and beliefs, although they may embody representations which operate in a similar way.

The situation is particularly complicated in the case of Freud's agencies. For while the ego and super-ego are clearly meant to be functional neural systems, teleologically described, these systems are also understood as embodying neural prototypes derived from actual persons in the environment. Freud's mode of explanation combines the idea of functional systems with the observation that the way persons actually function depends upon the prototypes by means of which they represent themselves and their relations to others. The ego thus embodies the prototypes which the child forms of the parents in their role as agents acting to satisfy desires, and the super-ego still earlier and cruder prototypes ('the earliest parental imagos') of regulating and controlling figures, laid down in relation to the infant's own basic impulses, such as those to eat and defecate, and severely distorted by early emotion and projection. Hence Freud describes the operation of these systems in terms of the motives of the basic prototypes which the systems embody. The super-ego is thus, e.g., 'an agency...which observes and threatens to punish' and which in some cases of disturbance becomes 'sharply divided from [the] ego and mistakenly displaced into external reality.' (XXXI, 59, 64)

Such descriptions are likely to seem at once mistakenly abstract and anthropomorphic; but in fact they serve relatively precisely to generalize over clinical data. Take again the example of the Rat Man's cowering in fear of punishment from Freud, while recovering memories of his father as actually punishing him, and seeming just such a terrifying figure as he had been taking Freud to be (xxxiii - xxxvi). This is one of many cases which fits Freud's description above: a part of the ego, which observes and threatens to punish, is here seen to be split off and displaced into the external world (in this case into the figure of the analyst.) This part, in turn, is apparently related to a distorted prototype of the patient's punishing father, as was emerging in conscious memory. And there is reason to suppose that the activation of a similar prototype -- in the encounter with the Captain who told him of the rat torture, and who really was fond of cruelty and physical punishment -- served to percipitate his breakdown in the first place.

So Freud and Davidson divide the self in different ways. Freud postulates partly distinct agencies which we describe in terms of figures with desires and beliefs, but which are ascribed these in what is ultimately a metaphorical way; and Davidson postulates partly distinct groups of desires and beliefs, which we may describe as agents, but only metaphorically. Johnston, by contrast, discusses the same topic in terms of the postulation of 'primary homunculi', which are real agents, with real motives (see his footnote 25). These are, therefore, clearly not the same as Freud's real agencies with only metaphorical motives, or Davidson's real motives which are only metaphorical agents. (Johnston also considers explanations which take 'the intentional stance' to things like plants, but these are not representation-ascribing teleological explanations at all.)

Since Johnston's criticisms are directed at a conception of division distinct from that of both Freud and Davidson, they apply to neither. This is clear from Johnston's own account of his critique. After describing 'a homuncularism which solves all the paradoxes of self-deception we have encountered' Johnston urges that

> This account can be discredited so long as we do not allow its advocates the luxury of hovering non-committally between the horns of a dilemma: either take the subsystem account literally, in which case it implausibly represents the ordinary self-deceiver as a victim of something like a multiple personality, or take it as a metaphor, in which case it provides no way to evade the paradoxes while maintaining that intentional acts constitute self-deception... (82)

Johnston's arguments enlarge on this claim, and also enforce his earlier remarks at pp 64-5 about the 'puzzles' inherent in homuncular explanation. But while an account of division in terms of 'primary homunculi' might well be impaled on this dilemma, those of Freud and Davidson clearly avoid both horns, and in different ways; so neither is in the least discredited. Freud's distinct systems constitute one person, and the way these are ascribed motives, although metaphorical, is nonetheless genuinely explanatory, as far as it goes. Davidson's 'constellations' contain genuine motives, which serve to explain in the usual intentional way, except that their working is in some respects abridged; and they are not distinct agents at all. So in neither case is there a threat of multiple personality, nor of the substitution for intentional explanation of unacceptable metaphor. Davidson can avoid the paradoxes by reference to cohesive motives and intentional acts, and Freud by reference to agencies and their goals; and these kinds of explanation are coherent, both internally and with one another.

We can make this clearer by starting with Davidson's account of akrasia, and moving from it towards the kind of description in terms of the super-ego which we have already considered. Davidson gives the following example, taken from a note in Freud's case history of the Rat Man:

> A man walking in a park stumbles on a branch in the path. Thinking the branch may endanger others, he picks it up and throws it in a hedge beside the path. On his way home it occurs to him that he branch may be projecting from the hedge and so still bge a threat to unwary walkers. He gets off the tram he is on, returns to the park, and restores the branch to its original position...It is easy to imagine that [he] realizes that his action is not sensible. He has a motive for removing the stick, namely that it may endanger a passer-by. But he also has a motive for not returning, which is the time and trouble it costs. In his own judgment, the latter consideration outweighs the former; yet he acts on the former. In short, he goes against his own best judgment.

Davidson explains how this example fits his account of akrasia, and also indicates how it might be deepened, through exploring his conception of a divided mind.

> ...Recall the analysis of akrasia. There I mentioned no partitioning of the mind because the analysis was at that point more descriptive than explanatory. But the way would be cleared for explanation if we were to to suppose two semi-autonomous departments of the mind, one that finds a certain course of action to be, all things considered, best, and another that prompts another course of action. On each side, the side of sober judgment and the side of incontinent intent and action, there is a supporting structure of reasons, of interlocking beliefs, expectations, assumptions, attitudes and desires. To set the scene this way still leaves much unexplained, for we want to know why this double structue developed, how it accounts for the action taken, and also, no doubt, its psychic consequences and cure. What I stress here is that the partitioned mind leaves the field open to such further explanations...

To think about further explanations it will be useful to replace Davidson's example from Freud with another to which it is closely related. In the original example it was plain that the branch was more dangerous in its original position, so that the incontinent intent was also hostile. In this Freud took the example to be similar to many from the Rat Man's own behaviour. Thus once when his lady was leaving,

> [The Rat Man] found a stone lying in the roadway and had a phantasy that her carriage might hit up against it and she might come to grief. He therefore put it out of the way, but twenty minutes later it occurred to him that this was absurd and he went back in order to replace the stone it its position. (X 307)

It is easy to imagine that the Rat Man also thought it would be best, all things considered, to let the stone remain in the safe place to which he had moved it, so that his action in moving it again was akratic. Here, however, we know something further about 'two semi-autonomous departments of the mind' each with many co-operating motives, one of which was for, and the other against, the akratic act. The Rat Man's attitude towards his lady was marked by the same deep ambivalence and conflict as that towards his father, as shown in the fact that he also frequently wished the rats on her, and suffered as a result -- particularly, it seems, when she vexed him by doing things like going away from him, as in the example above. On the side of moving the stone again, then, were arrayed a group of motives hostile to the lady, and shown also in the original phantasy that she might come to grief on it; while good sense (as well, perhaps, as the constellation involving the desire to protect her in accord with which he first moved the stone) would council letting it lie, in its safe new place.

The Rat Man was often ready to acknowledge his 'vindictive impulses' (X 185) toward his lady, so the episode as described might not have involved even ordinary self-deception. Still we can easily imagine that it did, and that this can be explained in Davidson's way. Suppose, having told Freud of the episode, the Rat Man had tried to explain his bothering to return -- to put the rock in what had originally struck him as a dangerous position -- by saying 'I did it to adhere to the ideal of rationality.' There might be a good deal behind this: as he walked along the road, say, he muttered rhythmically to himself 'I must be rational, I shall be rational, I shall leave things as they were.' And acting rationally, we might suppose, was one of his ideals, and one he took Freud and himself to share, but which he knew he often fell short of.

We can imagine that this made him more comfortable, but that it was self-deception. For he recalls, say, that -- although he was scarcely aware of it, and did not think much about it at the time -- he was in fact feeling angrier with his lady with each step he took along the road, and hence with each rhythmic muttering about doing what was rational; and as he walked away he imagined her carriage smashed to bits on the stone he had put back, and thought ' That will serve her right, for having dared to abandon me!' Material like this might lead us to judge that he had moved the stone as a result of an impulse to harm his lady, but that (in accord with a cohesive constellation of motives aimed at) wanting to look better in his own eyes, he had made himself think that he was doing it to adhere to an ideal of rationality, and had done so partly by talking to himself.

Such an explanation would presuppose that the agent's motives fell into groups partly comparable to those of a deceiver and a deceived, and also that he was not fully aware of their operation in these constellations at the time. Ascribing this lack of connecting awareness, however, would not be treating him as two distinct agents: for the motives were all his, and he was more or less aware of the thoughts and feelings connected with their operation, and clearly relevant to their ascription, all the time. Hence also he might readily acknowledge in retrospect that he had been deceiving himself with his talk of rationality, and that this had involved something like the 'flight from anxiety' which Johnston emphasizes (85); and in this he might also appreciate that the 'protective' constellation of motives on which he had acted in stressing rationality was a natural concomitant of his ambivalence, which required him to be more or less unaware of his real motives so as to act in accord with his hatred without suffering guilt. And we might know, and he might be able to acknowledge, that in deceiving himself in this way he had chosen his means well, because he knew his man: the line he fed himself worked because it was flattering, and it was one which he could be depended on to fall for.

Some divided constellations of motives thus are, or might be, clear enough; and these might serve to explain instances of akrasia and self-deception in considerable detail. Still, as Davidson says, this leaves much unexplained: for we want to know about the causes of these divisions, and where relevant their cure. This is what Freud's account tells us. Division or fragmentation in the self goes with division or fragmentation in (the representation of) those to whom the self has been most fundamentally related. The 'double structure' to which Davidson refers can be a structure of love and hate, ultimately built around disparate 'imagos' or prototypes of the parents, which were the earliest objects of these emotions, and hence the objects towards which they were directed in their most primitive forms. These early prototypes remain active in us, and shape our representations of ourselves and of subsequent objects of thought and feeling, and so partly determine the thoughts and feelings themselves. In this they contribute to the forming of what Freud called the ego and super-ego. Thus the image of a prohibiting and punitive father can both cause a rebellious and resentful desire to punish that father in return, and also be incorporated in a conscience whose punitive severity renders its possessor liable to suicidal depression. Such images can emerge in analysis, as indicated in the introduction, and hence be modified by further experience and thought. Still, the altering of basic psychic prototypes is a far more difficult matter than unravelling a piece of self-deception. In particular, such change requires what Freud called working through, and this can be a time-consuming process; but the psychic consequences are accordingly more far reaching.

The understanding of such early prototypes, and hence of division in the mind more generally, was significantly advanced by the work of analysts who followed Freud, and in particular by Melanie Klein. She was able to extend Freud's methods to the analysis of even very young children, by allowing them free and unhibited play as well as free association; and she observed that in these conditions children not only played out the satisfaction of the childhood desires which Freud had hypothesized, but also others, which were more extreme, and which allowed her to extend his theories in a systematic way (cf footnotes 27,29, and 22 in the introduction). She found that very disparate images of the parents, and hence of the self, seemed operative in disturbed children from early in life; and that the earliest prototypes, which lay at the root of all others, were the most fragmented, incoherent, and extreme of all. Hence she hypothesized that the original conflict-engendering images were laid down in early infancy, before the baby developed a working grasp of the concept of identity. The infant liable to violent emotion and excessive projection could not, in Hume's phrase, 'unite the broken appearances' of the parents, by synthesizing them into coherent wholes; and this incoherence was reflected back in its infantile experience and image of itself. Thus on Klein's account a fundamental task of infancy is that of integrating our experience, including that of persons, by means of the concept of an enduring and spatio-temporally coherent object. Psychoanalysis traces divisions in the self to failures in this original synthesis, and so provides an explanation of the 'double structure' which is both conceptually and empirically deep. And in the case of the self, the understanding of the broken images which are the causes of its division tends also to knit these images together, and so to provide the means of its cure.