PERSONS

A Study in Philosophical Psychology

RAZIEL ABELSON

Professor of Philosophy
New York University

ST. MARTIN'S PRESS NEW YORK
agents, as authors of their own actions, avowals, goals and reasons. Following Strawson’s analysis of the concept of a person as indefinable in terms of necessary and sufficient properties, and defending this insight against A. J. Ayer’s criticisms, I suggest that the concept of a person is a status-concept, analogous to more specific status concepts such as governor, judge, doctor, spouse and friend, in fact, constituting the genus of which these social roles are species. I then argue for the importance of distinguishing the concept of ‘self’ from that of ‘person’, and suggest that the concept of self has two uses, one of which is the trivial grammatical role of self-reference, while the more philosophically interesting use is that explained by Sartre and Freud as a projection of one’s future goals.

The analysis of the meaning of ‘self’ is applied, in Chapter 7, to the resolution of the paradoxes of self-deception. Four attempts to resolve these paradoxes are discussed and shown to be inadequate, namely Freud’s multiple self model, Sartre’s dialectical model, the translational theories of several analytical philosophers (Demos, Siegler, Gustafson and Penelhum – adequately criticised by H. Finnegar, and finally, Finnegar’s self-anputation model. I then offer a synthesis of the partial insights of these four approaches, in terms of a distinction between first and second order acts of self-deception which, in turn, centres on the Melden-Danto distinction, introduced in Chapter 3, between basic and complex actions. Although the paradoxes of self-deception appear, on this new account, to be resolvable, it turns out that one piece of the puzzle still eludes us. The last piece is provided, but only in outline, since we find that it can only be fleshed out by the avowals of the agent who possesses adequate self-knowledge. In the final chapter, the philosophical sense of ‘self’ explored in Chapters 6 and 7 is applied to the definition of ‘self-knowledge’ and ‘self-interest’, concepts that are shown to presuppose a community of moral agents whose avowals of their self-interests constitute prima facie reasons for action. It follows that morality and self-interest converge, and this fact provides the answer to the question ‘Why should we be moral?’ Thus the concept of a person entails that of a moral community of autonomous agents, each of whom recognises the rights and interests of all others.

The path I have followed in this work was lighted by Ludwig Wittgenstein’s Philosophical Investigations, although the first road-map was drawn by Ryle’s Concept of Mind. I have acknowledged my debt to Ryle in references throughout this book, but since Wittgenstein refrained from theoretical generalisations, I cannot claim his authority for the theory of persons offered here. I have thought it appropriate, therefore, to indicate my debt to Wittgenstein by beginning each chapter with a remark from his writings and conversations that has suggested to me the view I have taken in the chapter.
PSYCHOLOGICAL LANGUAGE

Why do words like ‘believe’, ‘motive’, ‘anger’ and ‘feeling’ belong on a list of psychological terms, while ‘round’, ‘mutation’, ‘magnetic’ and ‘electronic’ do not? There may be many reasons, but I would like to consider three that have been emphasised in recent philosophical discussions. These reasons are: (1) that psychological terms are those which a person can apply to himself without observation, (2) that their correct application essentially involves normative judgment, and (3) that they are intentional concepts.

Non-observational self-ascription

G. E. M. Anscombe, P. F. Strawson and Sidney Shoemaker have suggested that the distinguishing feature of psychological language is that it can be applied by a speaker to himself without prior observation of what he, the speaker, is doing, while it can be applied to others only by observing their behaviour. I can say truly that Smith is suffering pain only after watching Smith’s facial expressions and actions, or hearing him groan or say that he is in pain, but I can say truly that I am suffering pain without watching myself in a mirror, or listening to my own voice. In contrast, I cannot know that I am moving through space with a certain velocity, or that I weigh 150 pounds, or that my body is composed of cells, without observing my body and its relation to other objects. We may sum up the differences by saying that in psychology, self-knowledge has privileged authority over knowledge of others, while in natural science, all knowledge rests equally on controlled observation. Nineteenth-century introspectionist psychologists like Wundt and Titchener correctly noted this fact, but they drew two unfortunate conclusions: (1) that self-knowledge is indubitiable, a conclusion which is clearly false, and (2) that self-knowledge involves introspection or inward observation of a ghostly realm of mental events. This obscure explanation of the clear fact that each person has privileged knowledge of himself was rightly abandoned by behavioural psychologists who unfortunately threw out the baby with the bath water. In rejecting the obscure doctrine of introspection, they also closed their eyes to the very important fact that psychological beings do have non-observational self-knowledge, and they held that psychology is exclusively knowledge of the observational behaviour of others, relegating self-knowledge to the humiliating status of ‘verbal behaviour’ (a term made current by B. F. Skinner). The well-known joke about the psychologist who meets a colleague and asks politely, ‘How am I?’, is a reductio ad absurdum of the notion that self-descriptions are conditioned responses symptomatic of the state of the speaker, but no more worthy of belief than the barking of a dog.

Conceptual Dualism

One need only note the dependence of psychological inquiry on questionnaires, interviews and other types of self-reports to see that the reliability of most of our self-descriptions is an essential condition for the very existence of a science of psychology. If subjects could not tell us what they see, there would be no science of perception; if they could not tell us how they feel and what they want, there would be no knowledge of motivation; if they could not tell us what they believe and what they dream, there would be no psychiatry. Behaviourists might reply to this argument by pointing out that they have discovered a good deal of information about animal psychology even though animals cannot talk about themselves. But it seems doubtful that we could know anything about the psychological states of animals if it were not for their resemblances to human beings who do talk about themselves. Animal behaviour sufficiently resembles certain stereotyped patterns of human behaviour so that we feel justified in describing animals in the same terms by which we describe ourselves— as angry, depressed, fearful, hungry, intelligent, and the like. But the obvious limitations on this kind of analogical reasoning from our own states to those of animals have led behavioural psychologists to take refuge in technical jargon which, for some inexplicable reason, they still call ‘psychology’. For if the technical vocabulary of ‘reinforcement’, ‘stimulus-response’, ‘tension reduction’, etc. is detached from the language of non-observational self-description that gives it its psychological content, it becomes hard to see what there is for psychology to explain that physiology and physics cannot take care of without its help.

Normative judgment

A second peculiarity of psychological language that distinguishes it from the language of natural science was emphasised by Richard Peters in his monograph, The Concept of Motivation, and Peter Winch, in The Idea of a Social Science, although it was first suggested by Gilbert Ryle in The Concept of Mind. Psychological concepts apply only to creatures that exhibit certain skills. ‘Perceives’, ‘knows’, ‘responds’, ‘avoids’ are descriptions that imply some kind of ability even in failure. A rat can neither pass nor flunk a philosophy test, a guided missile can no more misjudge the position of its target than it can correctly perceive it. Psychological descriptions are thus more than mere descriptions; they are also evaluations of behaviour. Suppose, for example, that a psychologist is studying the emotional behaviour of dogs. He gives a dog an electric shock, notes that the dog jumps away from the electrode, and describes the dog’s response as ‘fear’. Yet motion away from an object is not by itself an adequate criterion of fear. A torpedo repelled by a magnetic field can hardly be described as frightened by it. Moreover, a dog will move away from objects for
motives other than fear – he will avoid objects that bore him or displease him, and even things that attract him, as in the case of the canine coquette who teases her mate by running away. In order to identify correctly the particular emotion that motivates the observed response, as fear, boredom, displeasure or coquetry, the psychologist must first evaluate the stimulus as fearful, or boring, or unpleasant or tantalising. But such evaluation makes use of normative standards that happen to be the conventional standards of the psychologist’s society, standards which he projects into the poor dog who is in no position to complain about the relativity of normative standards. Suppose the dog were clever enough to talk, and suppose he regarded electric shocks as rather thrilling: he might advise the psychologist to read Jean-Paul Sartre before imposing his bourgeois values on canine bohemians.

The essential role of normative judgment in psychological description makes the deductive model of explanation employed in natural science inappropriate to psychology. By the deductive model I mean that an event is explained by deducing it from a set of laws together with antecedent conditions that we call the ‘cause’. Now any so-called law that might be used as a premise in deducing a psychological event would have to be phrased in terms that presuppose normative judgments. For example, take the behavioural law: ‘Whenever an animal is confronted by a fearful stimulus, it tries to avoid that stimulus.’ How are we to determine objectively whether a stimulus is fearful? What is fearful to a coward is not so to a hero; a penthouse apartment is awesome to an acrophobe, but delightful to a social climber. The sight of a rare insect inspires terror in the insectophobe and joy in the entomologist. The behaviourist might argue that these cultural and idiosyncratic differences of normative standards among human beings do not apply to animals, and that for this very reason, animal psychology can be studied with greater scientific objectivity than human psychology. But the degree to which this is true is also the degree to which it is pointless to speak of animal psychology at all, because psychological predicates like ‘fear’ no longer do any informative work. It may be true that all dogs are repelled by just the same stimulus objects. If so, then we could formulate laws like: ‘All canines are repelled by electrically charged objects.’ But notice that the word ‘fear’ has now dropped out of the law, since it serves no informative purpose. The reason why we distinguish motivating states like fear, anger, boredom, etc. in describing human behaviour is precisely because human beings do not always respond in the same way to the same stimuli. Concepts like ‘fear’, ‘anger’, and ‘boredom’ serve to explain these differences of response by indicating the reason why a person avoids an object at one time and embraces it at another, and such reasons have essential reference to the normative standards of the agent. Where standards are so stereotyped that all agents respond to a given stimulus in the same way (as in the case of avoiding a red-hot stove), psychological intervening variables become vacuous and can be dispensed with altogether. If all human responses were like the automatic response to a hot stove, there would be no need for language that distinguishes human fear from magnetic repulsion or romantic love from gravitational attraction.

The difficulty here has to do with the subtle difference between observable properties of things and criteria of normative judgment. Let us consider this difference briefly:

Any observable property can serve as a criterion or standard of normative judgment, as a result of which it is easy to confuse the mere presence of the property with its role as a normative standard. For example, a poor man like myself uses mileage and durability as standards for preferring a car or judging it to be good. But the fact that a car is good according to these standards is a fact of a different kind from the fact that it is economical to run. A rich man may regard the car as bad for the very reason the poor man regards it as good, according to Veblen’s theory of conspicuous consumption. Thus for the poor man to say that the car is good is not equivalent to saying that it is economical to run. For the rich man can agree with him that the car is economical without agreeing that it is good. For the same reason, it is one thing to report that a dog moved away from an electrified object and quite another to say that the dog avoided the object, or that he manifested fear of it; the latter description implies that the dog’s response was appropriate to a fearful stimulus, which in turn implies that the stimulus object is in some way bad, or undesirable. Now normative standards are often so conventional and stereotyped that we take the normative judgment as tantamount to an empirical description as when I, as one poor man, say to another poor man that a certain car is a good one. He naturally understands me to be informing him that the car in question is economical and durable. Similarly, when a psychologist reports that a dog feared an electrified object, we take that as tantamount to saying that the dog jumped away from the object, because standards of fearfulness of stimuli like electric shocks and standards of appropriate response are stereotyped among humans as well as dogs. So if we confuse values with empirical facts in such cases, no practical damage is done. But this is true only of very rudimentary situations. Consider, for example, that a boxer often avoids a blow, not by running away from it, but by moving toward it and clinching, which is not a standard fear response. Again, if a person knows that an electric shock will cure him of an illness, or if he has formed an eccentric fascination for electric shocks, the difference between value judgment and empirical fact becomes crucial, since he might embrace the electrode in fear, or move away from it in fascination. If there were canine saints and canine masochists, capable of abstaining from food no matter how hungry they were, the differences
between the motives of hunger, duty and self-punishment would be essential to our descriptions of their food behaviour. But if such things as canine saints and canine masochists are excluded as possibilities, then canine psychology is indistinguishable from canine physiology, and psychological descriptions are superfluous restatements of physiological descriptions.

**Intentionality**

I would like now to consider a third difference between psychological language and physical language, and then I shall try to tie these three distinctions together. Some philosophers have defined psychological terms as those possessing the property of intentionality, where by ‘intentionality’ they mean a relation between a subject X and an object Y that holds independently of the existence of Y. Thinking, believing, loving, wanting are intentional states in the sense that X can think about, believe in, love or want Y whether Y exists or not. Franz Brentano was the first to emphasise this feature of psychological language, and Roderick Chisholm has made effective use of it in an illuminating critique of behaviouristic psychology. 5

This peculiar feature of psychological language helps further to explain why psychological inquiry cannot be modelled on inquiry in the natural sciences. If the object of an intentional state like belief or emotion need not exist, then it is impossible to verify sentences about such states by means of observable evidence alone, and it is equally impossible to formulate empirical laws associating such states with observable behaviour. For example, we cannot say that an organism always avoids an objectively fearful stimulus, but only that an organism always avoids what it believes to be a fearful stimulus, for there may be no external stimulus at all, as in the case of hallucination, or the real stimulus object may be innocuous, as when an animal is frightened by a shadow. Thus concepts like ‘fear’ and ‘avoidance’ are intentionally related to objects independently of the real properties and even of the existence of such objects and cannot therefore be operationally defined in terms of observable events alone.

The criterion of intentionality employed above has been criticised for begging questions about the meaning of ‘existence’ and ‘reality’, and for being both too broad and too narrow to cover the gamut of psychological discourse. Many plausible counter examples have been offered by W. Sellars, D. Rosenthal, H. Morick, and others. However, whatever the deficiencies of the Brentano-Chisholm criterion, they need not affect our discussion, since all that matters here is that the truth value of some ‘crucial’ (in a sense explained in Chapter 2) psychological statements be dependent on the beliefs of, or the descriptions acceptable to, the subject of the statement in question. For example (thanks to Fred Dretske), 'The child saw his father take a stick out of the closet', when used intentionally, becomes false when 'umbrella' is substituted for 'stick', but remains true when used non-intentionally. I am of course employing 'intentional' as a feature of our use of sentences, and not as a semantic property of sentences themselves. Perhaps I should say, more precisely, that I am stipulating this meaning of the technical term 'intentionality' for my own purposes.

**PSYCHOLOGICAL EXPLANATION**

We have considered three features of psychological language that distinguish it from the language of natural science: non-observational self-application, essential reference to normative standards, and intentionality. These three peculiarities have an important bearing on the kind of explanations appropriate to psychology and other behavioural sciences, like sociology, anthropology and history. I want to argue now that explanations of human actions — as contrasted with explanations of organic inability to act — are of a fundamentally different type from explanations of natural phenomena in physics, chemistry and biology.

It is fashionable these days to take the view that all explanations are of the same general nature, a view that was most thoroughly articulated by Ernest Nagel in his monumental study, The Structure of Science. 6 Logical empiricists, whose views have been accepted as canonical by many American psychologist, grant that there is a difference between deterministic and probabilistic explanation, that is, between explaining why an event had to happen and why it was likely to happen, the difference lying in our inability to isolate all the causal factors that determine the event in question. But they recognise no other kinds of explanation than these. According to this view, explanation and prediction have the same logical structure and differ only with respect to whether the event in question is past or future. To explain is to retrodict, to predict is to explain in advance. To do either is, essentially, to deduce an event or its probability from antecedent conditions by means of linking generalisations. If the generalisations are statistical, then the explanation is probabilistic; if the generalisations are laws in the sense that they cover all cases, then the explanation is deterministic or strictly causal.

Now it is a notorious fact that there are no non-trivial deterministic laws of psychology or of any other behavioural science. There are, of course, plenty of valuable statistical generalisations about behaviour that help us to predict the probability of individual or group behaviour under well-defined antecedent conditions, but these statistical generalisations are descriptive rather than explanatory. For example, the generalisation that most three-year-old children have temper tantrums enables us to predict that Johnny will (probably) have temper tantrums,
is scientifically pointless to construct theories of motivation, for the scientific value of a theory lies in the new laws that can be deduced from the theory. I suggest, therefore, that all causal theories of motivation are products of conceptual confusion between causal and rational explanation. In the remainder of this chapter, I shall try to indicate why psychologists are so often tempted to confuse these two kinds of explanations, and why it is important for them to resist the temptation.

Psychoanalytic theorists sometimes describe human motivation on analogy with an iceberg—onethenth above the surface and nine-tenths below. But it can be argued on philosophical grounds that the inverse of this analogy is more apt: motivation must be nine-tenths conscious or unconscious, and only one-tenth or so clandestine and obscure. For, as Sidney Shoemaker has argued in Self Knowledge and Self Identity, we could not possibly learn to use the concepts of reason, desire, feeling and intention if this inverse analogy did not hold. We learn to use these concepts primarily in a self-ascriptive way. Piaget describes how small children first employ phrases like ‘I want X’ as commands or requests, long before they can talk in a purely factual way about what other people want. Thus in order to understand the logic of ‘want’, we must take note of the fact that ‘want’ is primarily self-ascriptive, and only derivatively descriptive of others. As self-ascriptive, there are no observable criteria of its application—I need not look at myself in a mirror in order to say ‘I want a cup of coffee’. Nor, in ordinary circumstances, need the waitress scrutinise me to see if I really do want coffee. (By ‘ordinary circumstances’ I mean I am not drunk and I don’t look like a hobo or a madman.)

Now, if the primary use of psychological predicates—the use that we learn first—is non-observationally self-ascriptive, then it follows that, under normal circumstances, we must necessarily take a person’s description of himself at face value. When I tell a waitress that I want a cup of coffee, she takes my word for it. Taking a person’s word for his psychological descriptions of himself is not a privilege we extend to him for good conduct, nor is it a certificate we award him when he has completed 300 hours of psychoanalysis. Rather, it is a logical condition for the meaningful use of psychological language. As Shoemaker puts it, psychological self-ascriptions are necessarily generally true, that is, they must be taken as true under standard conditions, where no extraordinary factors conducive to insincerity or self-deception can be found—factors such as delusion, intoxication, extreme fear, having something to gain by lying, etc. But if this is true, then the claim that the larger part of human motivation lies concealed beneath the surface of consciousness must be false, since it contradicts the very condition that distinguishes psychological language from physical language, namely, the condition of privileged authority for one’s self-ascriptions, which is the first thing the child learns from his parents about psycho-
logical language. Thus, in making the erroneous iceberg claim, psychoanalytic theory has tended to reduce psychological language to physical language. It is true of a physical iceberg that most of it lies beneath the surface of the water, and this is a fact that we can verify by diving underneath and looking. But there is no way to verify what lies beneath a man's consciousness except by so arranging conditions through psychotherapeutic procedures that the abnormal conditions for erroneous self-ascription are transformed into normal conditions for true self-ascription. Psychoanalysis can be regarded in this light as a technique for restoring the standard conditions of psychological self-description. In this respect, psychoanalysis is a more subtle and effective set of procedures, but of the same general type, as giving a drunk a cold shower, reassuring a frightened witness by palliating his fear of public or family criticism, or responding with kindness and patience to a hostile and suspicious child. A. C. MacIntyre has made this point persuasively, in his monograph, The Unconscious, in which he argues that, insofar as the concept of unconscious desires and motives has any verifiable sense, it is as an extension of the everyday notion of what we know about ourselves, not at the moment of action, but after calm reflection. Thus the techniques of psychoanalysis presuppose the very principle that psychoanalytic theory denies, namely, that normally we can trust what people say about their own feelings and motives.

But if so, then the difference between a conscious motive and an unconscious motive is not really like the difference between the visible part of an iceberg and the invisible part. It is more like that between knowing a poem that one can recite at will, and knowing a poem that one can piece together only after long and undisturbed efforts at recollection. Unconscious motives are the exception rather than the rule, and we can know they exist only if they become conscious motives when normal conditions of self-description are restored.

All this has an important bearing on the difference between causal explanation and psychological explanation. Freud tended to confuse the two because he thought in terms of something like the iceberg model, with the huge understructure of libidinous motivation constantly exerting a sinister influence on the agent's conduct and thought. Marx thought of social 'dynamics' in a similar fashion, with the substructure of class interests determining political policies, institutions and ideologies. And from these two men of genius we have inherited interesting but scientifically worthless theories of individual and social motivation, according to which, if we know the unconscious substructural forces and laws at work, we can deduce the actions of individuals or groups as a physicist deduces the path of a projectile. The reason why such theories are scientifically useless is that there is no way of identifying the substructural causes except by their effects, that is, by what the person or group says and does. But explanations of this kind are like

Molière's doctor who explained that opium puts one to sleep because it has a dormitive virtue. The explanation is vacuous because the alleged cause is simply a fancy name for the effect it is supposed to explain, and has no further predictive power. Its scientific 'cash value' is zero.

Psychological explanations in terms of reasons or motives are not causal explanations because they have no predictive power other than the mere projection to the future of behaviour patterns already observed in the past. It is natural then, that psychologists who want to regard their discipline as a science on a par with physics, should try to transform psychological explanations into causal explanations. But in addition to the vacuity of this enterprise, it also misses the point of psychological explanation. The value of an explanation in terms of reasons or motives is not that it enables us to predict anything, but that it enables us to evaluate the action in question. When I tell you that my reason for extending my arm outside the window of my car is that I was signalling for a left turn, you judge my action differently from the way you did when you thought I was merely exercising my muscles, or showing off by driving with one hand. Thus the very reason why a rational or psychological explanation is not predictively potent, namely that it is a re-description of the action it explains, is also what makes it extremely useful in another way; it reveals the purposive significance of the action from the point of view of the values of the agent and thus enables us to evaluate the action, together with its reasons, as good or bad, wise or foolish, successful or unsuccessful.

If someone slips on a banana skin, we do not ask for his motive. If I extend my left arm in order to signal for a turn, it is silly to ask what caused me to extend my arm. Causal explanation and motivation explanation are mutually exclusive. To visualise reasons and motives as psychic forces analogous to physical stresses and strains is to obscure the crucial difference between psychology and physics. Reasons enable us to judge if actions are right or wrong, intelligent or foolish. Causes enable us to judge that effects are neither right nor wrong, but involuntary and unavoidable.

Does all this entail that causal explanation has no role to play in psychology? If what I have said entailed such a consequence, it would hardly be worth a hearing. It is perfectly obvious that psychologists make effective use of causal knowledge in anatomy, biochemistry and neurology to explain abnormal behaviour and to restore the necessary conditions for effective human action. It cannot be denied that a large portion of psychology and psychiatry consists in the experimental discovery and application of knowledge of the organic causes of breakdowns and reseculations of the human animal. But this portion of psychology is psychology only in a rather negative sense; it is knowledge of what prevents effective action, rather than of what motivates
effective action. There are thus two kinds of psychotherapy: (1) the branch of medicine that identifies and restores the necessary organic conditions for human action and (2) the kind that tries to influence an already rational agent to promote his own and others' welfare by offering the agent good reasons for socially desirable conduct. It is worth noting, at this point, that Freud once defined psychoanalysis as 're-education'. Now education does not consist in instituting causal factors that guarantee effective action. A student who scored 100 on an exam without even studying, when and only when he swallowed a certain pill, would not deserve credit for scholarly excellence. The chemist who invented the pill would deserve whatever credit was due.

Where all the organic conditions for effective human activity are satisfied, behaviour can be influenced for better or for worse only by non-causal procedures such as moral exhortation, inspirational example, reward and punishment, or rational argument. The religious leader, the philosopher, the school teacher and the psychotherapist employ all these methods in varying proportions with a degree of success that depends more on their personal qualities than on their theological, metaphysical, educational or psychoanalytic theories.

2 Authority and Freedom to Avow

'Just try, in a real case, to doubt someone else's pain or fear' – Wittgenstein, _Philosophical Investigations_

"So you are saying that the word 'pain' really means crying?"
On the contrary: the verbal expression of pain replaces crying and does not describe it" – Ibid.

It is, I believe, a little-noticed fact that all psychological knowledge is erected upon a foundation of first-person, present-tense self-descriptions, felicitously named 'avowals' by Gilbert Ryle. The word 'avowals' is extremely apt because such assertions are not ordinary descriptions of states of affairs; they are semi-performative in nature, falling somewhere in between the pure performatives brought to light by Austin, such as 'I promise', 'I do thee wed' and 'I dub thee Sir Lancelot', and physical self-descriptions such as 'I weigh 150 pounds'. I intend to show that psychology cannot be a theoretical science like biology because its data are 'subjective reports' or avowals, and all apparently more objective psychological knowledge is built upon these data, thus sharing their subjectivity. Secondarily, I hope to shed new light on how and why the human animal has free will, despite the correct Hegelian–Marxist claim that the individual is a nexus of social relationships.

My general argument for the logical dependence of all psychological knowledge on avowals is this: There are no purely mental objects, states, events or processes, as Wittgenstein, Sartre and Ryle have sufficiently argued (further arguments are offered later, in Chapter 5). Consequently it cannot be the job of psychological statements to refer to and describe and nomologically explain such non-entities. Insofar as psychological statements refer to anything, they must refer (but indirectly and vaguely) to bodily states and processes or tendencies to such. The primary job of such statements is not referential or descriptive; it is to interpret and evaluate the states of agents, the actions toward which they incline, and the circumstances that determine the moral or practical significance of such states and actions. I want to suggest that the origin of evaluations of circumstances, states and actions is to be found in the normative judgments of the agent as expressed in his avowals. Without this foundation in avowals, I shall
maintain, more general psychological statements such as ‘Jones is neurotic’ or ‘Smith has leadership qualities’ would be unintelligible.

I will first try to clarify the peculiar logic of avowals and then develop the argument sketched above, in tracing their logical relations to psychological generalisations. In so doing, I hope to account for the four features that distinguish psychological discourse from natural science: (1) non-observational self-knowledge, (2) intentionality, (3) normative judgment and (4) rational rather than causal explanation.

TYPES OF SELF-DESCRIPTIONS AND THE ROLE OF AVOWALS

There are at least eight types of psychologically self-descriptive assertions, of which only two are pure avowals (in the sense of ‘avowal’ I shall try to spell out below). First I wish to try to clarify the connections and the differences between the six or so non-avowal types and the avowal types. All eight may be illustrated by the following sentences (including some sub-types):

1. I am an irritable person (a generalisation).
2. I was (or will be) shocked by his appearance.
3. I see (or saw) a shark in the water.
4. (a) I remember that there was a shark in the water yesterday.
   (b) I remember her face.
   (c) I dreamt of a golden castle with barred windows.
5. (a) I see something blue.
   (b) I see a bluish patch.
6. (a) I feel cold (or angry, or happy).
   (b) I believe in God (or that God exists).
7. (a) I want coffee.
   (b) I intend to get coffee.
8. I am signalling.

It hardly needs argument to achieve agreement that the first five types are importantly distinct from the last three, in that their truth conditions are radically different. Few would dispute that statements of types 1, 2, 3, 4(a) and 5(a) can be verified and falsified by publicly observable evidence, and that a person can assert them mistakenly even when he is sincere and is choosing his words carefully and deftly. Types 4(b), 4(c) and 5(b) are often claimed to be unfalsifiable in the above sense. Whether or not they are, depends, I suspect, on whether they are taken as weak claims to the truth of their propositional content, or considered rather as purely phenomenological reports. Let us take a moment to explore this complication:

Can ‘I remember her face’ be disconfirmed by inductive evidence? If I say this of a well-known actress and then describe her with woeful inaccuracy (as dark complexioned when in reality she is fair, as brown-eyed although her eyes are blue, etc.), then surely I can be said to be mistaken. It was not her face I remembered, but someone else’s (my mother’s, perhaps?) or perhaps I simply thought I remembered her but I didn’t.

But suppose I insist that nonetheless, by golly, I do remember her face, not someone else’s, even if my memory image of her face does not do naturalistic justice to its model. I maintain against all my critics that I have a distinct memory image of Miss X’s face, no matter how Picasso-like distorted it may be, and I can report that image accurately. For example, in reading about Lady Guinevere, I form an image of her and later remember that image even though there is no Lady Guinevere at all for my image to correspond to, if Quine is right versus Meinong. I can truthfully say that I remember the face of my imaginary Lady Guinevere. So long as I am willing to suspend the issue of the public reality of the object of my imagery, my sincere report of my memory image is not subject to challenge. In this sense, there are lots of things we remember whose public correlates may never have existed, or may have existed in a quite different form from the way we remember them. Think of scenes of your childhood, or Marcel Proust’s remembrances of things past, of the ghosts and goblins that remain as toys in our psychic attics (or should I say, basements?), figures in our dreams, fictional characters, etc. The ambiguity of realistic versus phenomenological memory reports was noted by Ryle in The Concept of Mind.

My point here is that, insofar as there is a phenomenological use of memory reports, the publicly incorrigible character of such reports may be accounted for in a way similar to that of avowals proper (see below), in terms of the authority and freedom of the subject to interpret and evaluate his own immediate state and inclinations. No one but the subject (provided he is clinically normal) has the right and the authority to decide what his experience means to him. This truism is, I think, semantically entailed by the very concepts of thought, desire, intention and image. So one’s so-called phenomenological memory report is less a report than it is an authoritative declaration, resembling, but not identical with, a jury’s verdict or an executive’s decree.

The same considerations hold of dream reports, and even more strongly so, because, as Malcolm noted there is no ambiguity, or, more accurately, much less ambiguity between realistic and phenomenalistic dream reports (whether, as Malcolm suggests, the former do not exist at all need not concern us here). The experience of remembering a dream and reporting it to oneself or to another is the only waking experience that provides knowledge of the occurrence and content of
the dream (setting aside the REM experiments which, as Malcolm shows, beg the very question they attempt to decide experimentally, for what have the psychologists to go on except the subject's report when they wake him up, and how can they check his dream time scale against their real physical watches?). One may be tempted as always to identify the conditions for knowing a fact (i.e. the evidence) with the fact itself, and so to conclude that a dream is just the experience of remembering a dream and that remembering it is identical with sincerely telling it. But such temptations should be resisted. As G. E. Moore noted long ago, in his ‘Refutation of Idealism’, the experience of phi must be distinguished from phi itself, or else the experience would not be an experience of something, and thus not an experience at all. Nor can an experience be of itself on pain of an infinite and vicious regress in defining the concept of experience. Thus the dream remembered and reported is not the remembrance or the report, but it does not follow from this point of logical grammar that a dream must be a real event that occurs in space and time. It may well be that, like mental images, ideas, and all other soi-disant ‘psychic phenomena’, dreams neither take place nor fail to take place, but are simply remembered and reported, just as after images neither exist on the wall nor fail to exist on the wall, but are the sort of things that are simply, as Ryle would say, ‘grammatical accusatives’ of a certain type of visual experience.

Considering now perceptual reports of types 3 and 5 (a and b), although I have said that perceptual reports are publicly falsifiable, a qualification must be made at this point, in the light of the ambiguities revealed earlier. We can use type 3 as a purely phenomenological report (e.g. all right, so there's no shark there, but I know what I saw, and what I saw was a shark with great big teeth, not a dolphin or a log). But such a use is rather infrequent just because it tends to mislead us without the kind of adequate contextual preparation we get, for example, in an optometrist's examination room. As for reports of type 5(a), as Austin has noted, they are ways of hedging against untrustworthy visual conditions, organic or environmental. 5(b) suffers from the ambiguity discussed above, between reporting seeing a real patch, and phenomenologically reporting an experience that may very well be hallucinatory, as when one rubs one's eyeball. Such reports, as Austin put it, are ways of "refusing to stick our necks out".

At this point, before we get to pure avowals, I would like to express some reservations about statements of type 6(b). Belief reports, like memory reports and certain perceptual reports, can perhaps be meant in either a strong or a weak sense, i.e. committally or non-committally. "I believe that the ice is thin", according to Ryle, can be falsified if the speaker shows no fear of skating on the ice. But this is true only if the statement is taken in its fullest practical import, as it usually would be, since belief statements are usually made in a context calling for decision and action. But surely this is not always so. There is no semantical or pragmatic anomaly in asserting 'I believe that parachute is safe although you'll never get me to jump with it'. We have a right to avail ourselves of a purely intellectual use of 'I believe' where no commitment to action is intended. The reason why language provides for such non-committal uses of belief declarations is that we sometimes want to know what people are willing to support intellectually (i.e. by evidence), independently of what they are willing to do about it, and this in turn is because willingness to act on a belief involves factors such as habits, skills, self-confidence and whatever hang-ups our psychic flesh is heir to.

AVOWALS PROPER

Now we can proceed to consider the differences between pure avowals of types 6 and 7 and the objectively corrigible self-descriptions of types 1 to 5.

The difference between the phenomenological uses of self-descriptions just considered, and the uses of pure avowals, is this: the former depend logically on realistic use; they are employed as hedging devices, i.e. as ways of avoiding responsibility for implications that follow from realistic claims in situations where, for organic or environmental reasons, one is not in a position to be sure of the truth of his claims. Pure avowals, on the other hand, are not in this way parasitic on realistic claims. Their role is to introduce (I am tempted to say 'create', but that would be too strong) new facts by declarational fiat, in somewhat the same way, though not quite the same way, that a jury makes a defendant a convicted man by pronouncing 'guilty', or a chief of state, as Stephen Toulmin noted makes a bill a law by signing it. Similarly, we pronounce what we experience to be painful or pleasant, we declare our beliefs and intentions, we adopt stances and postures by declaring 'I hate you' or 'I love you', we make decisions by declaring 'I'll have coffee', 'I vote for Smith', and we commend and grade by declaring 'I like X', 'I prefer Y'.

The avowal uses of language are closely related to, but not at all identical with the expressive and imperative uses. Wittgenstein mistakenly suggested, in his Investigations, that avowals are purely expressive, e.g. 'I feel pain' is a verbal substitute for a scream or a groan. But such a view fails to account for the truth content of avowals, as a result of which he falsely concluded that they are neither true nor false, thus not quite freeing himself from his positivist bondage. Now 'X feels pain' when said by Y and 'I feel pain' when said by X say the same thing, yet surely the former is not expressive and, since it is true
Persons

or false, so must the latter be. Equally, it would not do to equate ‘I want coffee’ with ‘Please give me coffee’, since ‘X wants coffee’ said by Y and ‘I want coffee’ when said by X say the same thing, yet the former is clearly not a command or request. Moreover, one can respond to ‘Give me coffee’ by saying ‘Nothing doing, nyet’; one can deny, disobey or refuse the command, but one cannot deny, disobey or refuse ‘I want coffee’; one can at best (or worst) disregard it, as waitresses so often do. Finally, the avowal makes essential reference to the speaker as subject, while the command or request does so only as direct or indirect object.

Nevertheless, it seems unlikely that avowals, like commands, would ever be employed were it not for their intended effect on the listener. We normally say, ‘I want coffee’ in order to get the waitress to give us coffee, or ‘I have pain’ in order to elicit either sympathy, an anodyne, or perhaps dental surgery. How then do avowals differ from commands, requests and exclamations?

As we have already seen, avowals, unlike their rivals, are true or false. They correspond or fail to correspond to equivalent third-person descriptions. ‘Jones has a toothache’ when said by Smith and ‘I have a toothache’ when said by Jones assert the same fact and are true or false together. On the other hand, ‘Give me coffee’ has no third person descriptive equivalent and is neither true nor false, nor is an expressive cry of pain.

Another difference, related to the above, is that avowals, unlike commands and expressive exclamations, have logical implications. There are semantical and syntactical rules permitting us to infer consequences from avowals. ‘I have pain’ permits us to infer that the speaker dislikes the state he is in and would, if he could, ceteris paribus, initiate some change in his circumstances. These differences naturally tempt us to lump avowals with physically comparable self-descriptions like ‘I weigh 150 pounds’. But that this temptation should be resisted is clear when we consider the following differences (aside from those already revealed above):

1. In a physical self-description, the first person pronoun ‘I’ is replaceable by the speaker’s name or any unique description of him without change of truth conditions. But this is not the case for genuine avowals. ‘I love you’ and ‘Raziel Abelson loves you’, even when said by the person known as Raziel Abelson, do not have quite the same modal conditions. They are not necessarily (although they may well be contingently) substitutable salva veritatis. For it is a contingent fact that my name is Raziel Abelson. I might have been substituted for another infant when I was born, or I might suffer from amnesia, or I may be pretending to be Abelson whom I have just murdered. Yet if I say sincerely ‘I love you’, it could not pos-

sibly be the case that I, whoever I really am, do not feel love for you. At least, it could not possibly be the case that my statement is false due to any misidentification of the subject of my statement, of the kind previously indicated.

2. As Anscombe and Strawson have pointed out, and shown the great importance of, we are in a privileged position to make true (when sincere) or false (when we have reason to lie) avowals without any need for observational evidence, but this is not the case for physical self-descriptions nor, as we shall see later, for general psychological self-descriptions. I know that I weigh 150 pounds on the basis of the same kind of observational evidence (namely, scale readings) by which anyone else could know this fact about me.

The similarities and differences noted between avowals and imperatives and expressive or performative exclamations, on the one hand, and physical self-descriptions on the other, tempt philosophers to assimilate avowals either to the former (Wittgenstein and Austin) or to the latter (classical empiricism and phenomenology). But to serve as data for psychological generalizations, avowals must be more objectively informative than imperatives, exclamations, and pure performatives, and they must also be more authoritatively incorrigible than physical self-descriptions or even general psychological self-descriptions. Thus to assimilate them to either side of the subjective-objective divide would deprive them of their distinctive value, which is just to bridge that divide by establishing facts about oneself that no one else can possibly establish (they can only guess, and could not even guess if there were not already the established practice of self-avowal).

One final point on the relation of avowals to performatives: In How to Do Things with Words, Austin assimilated avowals to performatives. This suggestion is, I think, as close to being right as it is possible to get while still being wrong. Avowals, like performatives, are unchallengeable by contrary evidence because they are not conclusions from evidence or matters of direct (introspective) observation. Thus it is as silly to say ‘You don’t really feel pain’ as to say, ‘You don’t really do’ (to a bridegroom saying ‘I do’) unless we mean only to bring attention to features of the situation that produce what Austin calls ‘infelicity’, namely, that the presupposed conditions that authorize the declaration are in some way at fault (e.g. that the priest at the wedding is really an escaped convict – and he is not Father Berrigan). But unlike performatives, avowals, even when challenged for good cause, such as signs of insincerity or of poor choice of language, remain authoritative, if not withdrawn voluntarily by the speaker and if insistently repeated (e.g. I assure you, I swear to you that I really do feel pain, all appearances to the contrary notwithstanding, and I assure you I mean pain and not just discomfort, etc.). We cannot deny the
speaker's authority for his avowals, without relegating him to the dustbin, if not of history, at least of the normal human community. To do this is to commit him to an institution for the mentally, morally or emotionally defective.

A further difference between avowals and performatives is that unlike avowals, performatives are not true or false statements, but ceremonialised verbal acts. In this respect, jury verdicts and official decrees are performatives, but not avowals. To say: 'I promise' is to do something namely, to promise. To say 'We find the defendant guilty' is to do something, namely, to convict him. But to say 'I feel pain' is surely not to feel pain, and to say 'I love Sally' is not in itself to love Sally, although it may help a bit to produce or intensify that love. The fact asserted is not identical with the act of asserting it, although they are interestingly linked, as we shall see in a moment. Thus avowals are challengeable in a way that performatives are not, namely, as insincere. (A note of caution is needed here about the ambiguity of 'insincerity'. An act of promising may be said to be insincere, but not in the same sense as that which applies to avowals. An insincere promise is one that the agent does not intend to keep, but this fact does not make it any the less a promise - it does not make it false to assert that he did indeed promise. Quite the contrary. But the insincerity of 'I love you' does make it false for the speaker to say what he said.)

Yet despite these important differences, avowals and performatives are closely related because they perform similar linguistic jobs of establishing facts by verbal means. In the case of performatives, the fact established is the utterance itself, whereas in the case of avowals, the fact avowed is not the avowal of it. Nonetheless, the avowal, like the performative, is logically connected with what it asserts, although the logical connection is considerably weaker. Under normal conditions, the sincere avowal is logically sufficient for its own truth, but not logically necessary, while a performative is logically both necessary and sufficient, not for its own truth, since it has no truth value, but for performing successfully the task for which it is designed, such as promising, marrying, authorising, etc., provided of course that the presupposed institutional conditions are satisfied. It is because of their logical sufficiency for what they assert that both types of utterances are unchallengeable as honestly false. And this logical sufficiency is their very function for the following reason:

We leave it to the agent what to make of the state he is in, whether to interpret it as love, as hatred, as pleasure, as pain, etc. We leave it up to him to decide these matters, as we leave it up to him to decide to promise, to choose, to authorise officially and to wed. Through avowals and performatives, one asserts his freedom and his rights as a person, and any decent society respects those rights and grants that freedom.

We grant these privileges to no other creatures on earth, not even to the very best robots. In this respect, Marx and Hegel were right, that the individual as a free agent is created by society, and Rousseau and Kierkegaard were wrong.

I have used the word 'privilege'. Are avowals and performatives really exercises of privileges? Clearly performatives are. One cannot promise or marry except within a framework of social conventions and mutual trust in which, like a credit card, one's privilege is respected by others unless and until he has so abused the privilege by dishonesty that it is taken away from him. But similarly we cannot avow pain or love or belief to behavioural psychologists who refuse us such privileged authority by taking our utterances as mere 'verbal data', to be explained but not to be believed, just as we ordinary humans would do with a robot.

Now why do we grant these cognitive privileges - why does our language provide for avowals and performatives? The reason lies, I think, in our respect for and our social power to grant individual freedom and authority, by which I mean the following:

We sometimes choose, act and react in ways that could not have been expected even on the basis of considerable knowledge of our past behaviour, habits, culture, etc. Our assumption that we can understand each other presupposes that we have values and goals in common but we also assume that each person arrives at his own unique balance among conflicts and priorities of values. Without this latter assumption, we need never consult him as to his wants and needs, nor need he consult us. Democracy is the institutional recognition that each individual has a unique value system, capable of creative variation at any moment. Democracy is thus the only mode of social organisation that reflects consistently the metaphysical-ethical concept of a person. To be a person is to have the right to normative authority, as Kant first noted. If we all aimed at the same goals to the same degree, like rats in a maze, there would be no need for avowals, as is the case in the society envisioned (naturally) by B. F. Skinner in Walden II. Like Kierkegaard vis-à-vis Hegel, I wouldn't want to be a robot resident of Skinner's behaviouristic paradise.

AVOWALS AND PSYCHOLOGICAL GENERALISATIONS

I will offer, first, an a priori argument to show that general psychological statements are logically dependent on avowals, in that their meaning must be unpacked partially in terms of avowals. Next, I will offer an illustration of the form that such explications might take.
The A-Priori Argument
Assuming, as it seems to me reasonable to do, that psychological language is indispensable for communicating facts that cannot be adequately described in physical language alone, it would seem to follow that psychological concepts must be defined in terms of avowals (not exclusively, but at least in part). For one of the features that makes psychological concepts irreducible to physical concepts is the fact that they can be self-ascribed without observation and this non-observational self-knowledge is precisely, as we have seen, what avowals are designed to convey. Any apparently psychological description of a computer or a robot can be translated without remainder into a physical description. For example, 'The computer became neurotic' translates into 'The computer had a short circuit', or, perhaps, 'A card was inserted involving division by zero, and the effect was circuit hyperactivity and breakdown'. 'The robot is angry' could be translated into some Rylean dispositional formula such as 'If anything approaches it, it will swing a hammer in that entity's direction'. The reason why psychological descriptions of human beings are indispensable (setting aside animal pets who are describable on analogy with humans) is that the human agent can avow his own states without observation and we normally take his word about them. Try, for example, doing social psychology or clinical psychology without avowal-eliciting questionnaires or therapeutic sessions. This is what makes the agent's psychological states irreducible to physical states, since the latter cannot be known without observation; no person or object is granted privileged authority about his physical states.

The Relation of Generalisations to Avowals
But just what is the logical relation between general psychological statements and avowals? It is to be expected that the exact relation would vary from one type of statement to another, depending on the psychological predicate involved, and also on possible ambiguities of sense, such as we noticed earlier with respect to realistic-phenomenological ambiguity, and also the ambiguity of the concepts of sincerity and belief.

Thus I will not attempt to explicate all psychological concepts, not even to provide a complete analysis of a single one. It should be enough, for the purpose of illustrating the logical relations I have referred to with admitted, but inescapable vagueness, to sketch a partial explication of a single psychological generalisation, whereby its logical dependence on avowals is brought to light.

Consider the statement 'John loves Jane'. What kinds of facts are entailed by this statement?

Note first that the explication of meaning in terms of what entails or is entailed by a statement is not the same as the question of evidence, i.e. of what facts, signs or symptoms lead us to conclude, or serve us to confirm that John indeed does love Jane. This latter question can be answered adequately by observing John's behaviour, but not so for the question of criteria, or meaning implications. Suppose that John is a deaf mute, unable to avow his feelings verbally, and ignorant of sign language. Suppose too that when he sees Jane, his eyes brighten, his lips become moist, and his pulse quickens. He seizes every opportunity to perform errands for her, strikes everyone who mistreats her, and looks depressed when she flirts with other men. It would be quite unreasonable to able to doubt that John loves Jane, despite his inability to avow his feelings.

So far, behaviourism seems correct in claiming that observable evidence is sufficient to establish and confirm the psychological state of love. Or, in a more sophisticated view like that of J. Fodor, at least the evidence establishes an overwhelming probability of the state of love as a hypothetico-construct like magnetic lines of force or gravitational attraction. Yet surely, if we are convinced that John loves Jane, as we have every right to be, on the basis of observable evidence alone in this particular case, this innocent fact in no way precludes that, if John were able to talk to a confidante, he would avow his love for Jane (and, were he not too shy, he would avow it to Jane herself). Now what if John learns sign language and communicates with a trusted friend, and denies to his friend in all sincerity that he loves Jane? We would probably conclude that John is nevertheless in love with Jane but doesn't realise it, represses it; in a word, deceives himself. Again, behaviourism and/or theoretical-hypothetical-construct psychology seems to win the day.

But now consider whether this case could conceivably be generalised as typical. Is it conceivable that we could have such adequate behavioural evidence of the state of love had we never established a language of avowals, or were we to abandon it in the future, as Richard Rorty has suggested we very well may? Could we, in such a case, still have a concept of love distinguishable from the concept of a disposition to behave in certain ways, striking out on some occasions, smiling on others, etc.? That is, would we have a concept of love as an emotional state of which a robot is incapable, and which is therefore not reducible to purely physical description? Note first that, if we did, it could still not be as rich a concept as it is now, for the very distinction employed above between love of which one is aware and unconscious love would make no sense at all, since the only criterion of awareness, namely, ability to avow one's state, would be unavailable. Nor, on further reflection, could we even have an impoverished concept of love that is any richer than a concept of a tendency, and thus there would be no point to the concept at all. For there would be
no way to distinguish the disposition, say, to do favours out of love from that of doing favours out of kindness, or the hope for future reward, or even inverted hostility, since we could not learn to distinguish these motives by comparing the behaviour of others with their avowals, and with our own behaviour and avowals when we are in states whose nature we can know and thus avow without self-observation. It might be countered that we can distinguish the various motives for patterns of behaviour by the goals toward which such behaviour is directed, but first, how could we distinguish the goal aimed at from that merely arrived at, without the help either of the agent’s avowals, or, assuming him to be like us, without our own avowable non-observational self-knowledge? (And we learn what our own state is when, as small children, we imitate the avowals of our parents.) And secondly, how many goals of human action are observable? Precious few, I should guess; in fact, we should all hope so, because we want to grow, and that means we raise our sights at new goals, and also because the higher, spiritual goals of man are not definable in the language of natural science.

INCORRIGIBILITY

The ‘incorrigibility thesis’, that full-fledged avowals are normally not subject to challenge (where by ‘normal’ is meant that the subject is not infantile, is in his right mind, knows the meanings of the words he uses and has no discernible motive to lie), has been severely challenged by science-oriented philosophers, most notably by Richard Rorty in a very influential essay, ‘Mind-Body Identity, Privacy and Categories’. It will be instructive to examine Rorty’s arguments to see how and why this challenge goes awry, even in the best of hands. Rorty maintains that psychological predicates like ‘pain’ are, in principle, eliminable from language and with them, the status of ‘privacy’ (i.e. incorrigibility) that distinguishes them from non-psychological predicates, in somewhat the way that references to witches and demons have been eliminated in favour of psychiatric diagnoses such as that of hallucinatory psychosis, and also in the way that we might someday cease to talk about tables and talk rather about clouds of molecules. His first analogy is nicely provocative, but it is the second that he stresses. Rorty observes that witches and demons are non-observable, explanatory entities that have been replaced by the explanatory entities of psychiatry, while tables are directly observable entities for which we have, as yet, no adequate observable replacements. Consequently “although we could in principle drop “table”, it would be monstrously inconvenient to do so, whereas it is both possible in principle and convenient in practice to drop “demon”. Rorty concludes that although it would be equally inconvenient to drop ‘X has pain’ in favour of the neurological statement ‘X’s C-fibres are stimulated’, the advance of neurology and encephalography may well result in our being able to observe directly that someone’s C-fibres are stimulated, and thus to drop references to pain as we have dropped references to demons, and could drop references to tables. The dispensability of psychological self-reports about pain, etc. would become clear when such self-reports are found to be overridden by their neurological equivalents so that their privacy or authority has been undermined. How would this come about? Rorty describes a test-case in the future, as follows:

The interesting case is the one in which... somebody (call him Jones) thinks he has no pain, but the encephalograph says that the brain-process correlated with pain did occur... suppose that Jones was not burned prior to the time that he hitches on the encephalograph, but now he is. When he is, the encephalograph says that the brain-process constantly correlated with pain-reports occurs in Jones’ brain. However, although he exhibits pain behaviour, Jones thinks that he does not feel pain [my italics, R.A.]. (But now as in the past, he both exhibits pain behaviour and thinks that he feels pain when he is frozen, stuck, struck, racked, etc.) Now is it that he does not know that pain covers what you feel when you are burned as well as what you feel when you are stuck, struck, etc.? ... The only device which would decide this question would be to establish a convention that anyone who sincerely denied that he felt a pain while exhibiting pain behaviour and being burned ipso facto did not understand how to use ‘pain’.

... We now see that the claim that ‘such a mistake is inconceivable’ is an ellipsis for the claim that a mistake, made by one who knows what pain is, is inconceivable... But when formulated in this way our infallibility about pains can be seen to be empty.16

Before I point out Rorty’s main mistake, I shall make two preliminary comments: (1) The eliminability of ‘table’ in favour of ‘cloud of molecules’ in no way helps to show that an entire type of observables is replaceable by theoretical expressions, since ‘cloud’ is as much the name of an observable as ‘table’. And if one tries to replace ‘cloud’ by some purely theoretical term, what will do the job? ‘Swarm’, ‘collection’, ‘bunch’, ‘congeries’, ‘crowd’? All these are names of observables, unless used metaphorically. Obviously as this point is, it seems not obvious enough to those who have such faith in theoretical language as to believe that our only reason for our non-technical vocabulary is ‘pragmatic convenience’. (2) Unlike ‘table’, ‘pain’ does an expressive and evaluative job, as Wittgenstein has pointed out.17
That is why ‘I have pain’ is incorrigible. When I say that I have pain, I do more than report a fact – I plead for assistance and express my abhorrence for the state I am in and my desire to escape it. If some day we were to employ expressions like ‘My C-fibres are stimulated’ to do this expressive-evaluative job, then such assertions made in the first-person present indicative would have been recruited to do the same job as ‘I have pain’, and would thus have acquired a psychological meaning (use). In such a case, ‘My C-fibres are stimulated’ would no longer function as a neurological report and it would no longer be incorrigible by public criteria such as encephalograph readings. When I then say, ‘My C-fibres are stimulated’, and you reply, ‘Not at all, the encephalograph shows no such indication’, I would be entitled to reply: ‘Never mind the encephalograph; I know when I want assistance, and when I don’t like the state I’m in!’ You then have two choices: you can disregard how I feel, concerning yourself only with the scientifically provable facts, in which case you have removed yourself from the moral community of persons who care about how others feel, or you will respond to me in just the way you would if I said that I have pain. The economy achieved by replacing ‘I have pain’ with ‘My C-fibres are stimulated’ would be merely terminological. The language game of authoritative avowals cannot be dispensed with, unless we are prepared to dispense with that democratic respect for the evaluations of persons on which the equality of individual rights and interests depends. As Wittgenstein and Strawson have put it, to accept the avowals of others as authoritative is to see them as persons, rather than as automatons.

The most serious mistake in Rorty’s argument consists in what appears at first sight to be a minor slip. In his account of his imaginary experiment of the future, in which an encephalograph reading overrides a pain-avowal, Rorty supposes that Jones exhibits pain behaviour, but denies sincerely that he feels pain. Now what exactly is pain behaviour? Presumably, groaning, twitching and attempting to escape what one takes to be the source of pain (in Rorty’s example, a hot object). Why do we call such gestures and actions ‘pain behaviour’? Is it not because, as Wittgenstein observed, they are natural expressions of pain, not merely contingent effects, but part of what ‘pain’ entails, so that someone who groans and twitches is letting us know that he is in pain? If he then says he is not in pain, either he is lying (which Rorty rules out ex hypothesi) or else of course he shows inadequate misunderstanding of the meaning of ‘pain’. But this has nothing to do with the encephalograph indication that his C-fibres are stimulated. It has rather to do with the inconsistency of his natural expressions of pain and his verbal expression. To claim incorrigibility of a pain avowal in such a case is to win an empty victory, since it is not the encephalograph that is overriding Jones’ denial of pain, it is his own gestural expressions of pain. The appropriate test-case for Rorty’s incorrigibility thesis would be one in which all of Jones’ expressions, both gestural and verbal, were consistent, e.g. he smiles, welcomes being burned again, and says he enjoys it, although the encephalograph registers that his C-fibres are firing. In that case we should say ‘So much the worse for the neurological theory that the firing of C-fibres is necessary and sufficient for pain.’

It was surely a slip on Rorty’s part to stack the cards in favour of his incorrigibility thesis by postulating pain behaviour in support of the encephalograph reading he wished to make authoritative. But it was a very revealing slip, for it shows how easily one can shift from psychologically significant behaviour to mere bodily movement, assume that nothing has been lost, and thus ‘prove’ that psychology is reducible to biology or physics. What we call ‘pain behaviour’ is not mere bodily movement; it is movement that expresses what we feel, and thus ‘tells’ others what we feel. In the way a groan or a twitch cannot be mistaken, but only insincere, ‘I have pain’ cannot be mistaken, but only insincere. It is true that ‘I have pain’ can be not only insincere, but also false, while a groan cannot be false, and this is so because ‘I have pain’ is both an expressive utterance and a report of a fact, the fact that I feel pain. Nevertheless, it is a report of a fact about which one cannot be honestly mistaken. It may well be that this fact can be explained neurologically, but the explanation cannot substitute for ‘I have pain’ because ‘I have pain’ is more than a pain report, it is also an expression of pain – thus itself a mode of pain behaviour.

The general thesis I have argued for is that avowals are our primary criteria for identifying each other’s immediate psychological states, and even our own long-term psychological tendencies (which are tendencies to get into the immediate states such as anger, pain, etc. that become known to others through our avowals primarily, and our behaviour secondarily). Behavioural criteria and symptoms, while often adequate within a conceptual framework based upon avowals, become hopelessly impoverished when deprived of that framework. And the reason for this is that avowals are our unique way of making public the states of which we have non-observational self-knowledge, the reason in turn for which is that human social evolution, in creating a language of avowals and institutions within which it can be effectively employed, has made us free and creative individuals.
3 Cause and Reason

"But why do you say that we felt a causal connexion? ... One might rather say, I feel that the letters are the reason why I read such-and-such. For if someone asks me "Why do you read such-and-such?" - I justify my reading by the letters which are there" - Wittgenstein, Philosophical Investigations

Having found, in avowals and performative acts, an intrusion into the natural world by human agents that must be incompletely determined, if its author is to have incorrigible authority, the question arises: does this entail that human actions are causally undetermined as well, and, if so, are they therefore inexplicable? Our answer to the first question will be affirmative and our answer to the second negative. I suspect that the aversion most philosophers have to indeterminism is due to their mistaken fear that it entails an affirmative answer to the second question. I shall try to allay that fear.

So far we have considered only verbal actions, and of these, only a subset: avowals and performative acts. We must now consider the logical relation, if any, between such speech acts and overt bodily actions in order to decide if the causal indeterminism of the former affects the latter. I believe it does.

I shall argue that voluntary human actions of any kind, verbal or non-verbal, are not explainable by causes, not in so far as they logically could have causes though in fact they lack them, but in the stronger sense that either to affirm causes of actions or to deny causes of actions is semantically anomalous, amounting to what Ryle has called a 'category mistake'. We have already considered why this claim holds for avowals and performative speech acts. I now want to show, with regard to non-verbal (voluntary) actions, that an adequate explanation of such actions involves avowals as part of the explainans, so that the indeterminism of the latter entails the indeterminism of the former.

The explanation of a voluntary action always, I maintain, involves reference to the agent's avowals because actions are explained in terms of reasons, and an agent's reasons can only be known - indeed, can only come into existence - through his avowals. Reasons are appeals by a litigant to premises of an argumentative conclusion, in justification or exculpation of what he says or does, and the litigant is the final authority, having authoritative non-observational self-knowledge, on

what his reasons are. It is for him to select the grounds on which he rests his case, grounds that manifest his choices of relative values and priorities, that is, of what for him constitutes adequate justification or exculpation. Insofar as we share his values we will find his reasons to be good reasons; to the degree that our own priorities differ from his, we may judge his reasons to be bad. If we differ too radically, we may be unable to see his reasons as even bad reasons, and we will consider him so irrational as to be unable to make authoritative avowals; in a word, insane.

The argument for this view leads us back to Ryle's Concept of Mind and the conceptual revolution which he and Wittgenstein began.

The Conceptual Revolution

Gilbert Ryle's The Concept of Mind charted a new direction for studies in philosophical psychology. Ryle's explicit aim was to demolish Cartesian dualism and its privileged access criterion of mind. In performing this demolition job, he also accomplished something positive. Intentionally or not, Ryle forged a new kind of dualism in the ashes of the old. In arguing that mind and matter are not different kinds of substances, Ryle revealed that mental and physical concepts are radically different kinds of concepts. Metaphysical dualism gave way to conceptual dualism.

But Ryle did not follow out the implications of his new dualism as far as they promise to lead. Others have done that since. Toulmin, Nowell-Smith, Hare, Mayo and Baier have deepened our understanding of the difference between normative and descriptive discourse, while Austin, Peters, Urmson, Anscombe, Hamlyn, Dray, Melden, and C. Taylor have discovered new categorial boundaries separating psychological concepts such as motivation, belief, perception, intention and volition from the language of natural events and processes. I think there is a good reason why the impact of The Concept of Mind on both ethics and psychology was equally great. The most important result of Ryle's attack on metaphysical dualism - a result which, for a reason that I shall try to explain below, Ryle himself did not fully appreciate - was the increasing realisation that the purposeful language in which we talk about human conduct is inescapably normative, and that psychology and sociology are logically dependent on ethics, rather than the other way round. R. Peters, in Motivation, and P. Winch, in The Idea of a Social Science, have brought these implications to light.

In his chapter on emotion in The Concept of Mind Ryle offered a dispositional analysis of affective states such as emotion and desire, according to which they turn out not to be internal states at all, but to be 'reasons' for predicting vague ranges of overt behaviour. Emotions
Persons

as motives, Ryle contended, are not inner springs of action; they are reasons for our acts, not causes of our acts. But Ryle failed to note the categorial gap between reasons and causes which his own analysis should have suggested to him. He assumed that reasons and causes worked together in the same type of explanation. The cause of an event or action is, for Ryle, some antecedent event or stimulus, while the 'reason' is a law-like conditional generalisation, licensing us to infer the occurrence of the effect or response. Ryle thus remained a scientific determinist on the issue of free action, while dissociating himself from mechanistic behaviourism and from mind-body dualism. What he failed to see, and what Peters, Dray and Melden have since brought to light, is that the concept of cause (in its natural scientific sense) cannot intelligibly be applied to the explanation of human action. For three centuries philosophers have been purging natural science of purposive language. The ill-fitting shoe is now on the other foot and we face the converse task of expunging pseudo-mechanistic concepts from the purposive language of human affairs.

CAUSES AND REASONS

Ryle began this task by distinguishing reasons and motives from causes. But his distinction did not cut deep enough. He recognised that psychological concepts are explanatory rather than referential, but he failed to see that a psychological explanation is not a causal explanation—that, in fact, the two are logically incompatible. To cite the motive for an act and to cite the cause of an event is, in either case, to answer the question 'Why', but the meaning of the question is different in the two cases. Dray, Peters and Melden have since revealed fundamental differences between psychological and causal explanation. They have argued that to ask 'Why' about a human action is to ask to make the action intelligible by filling out its purposive context, including the beliefs and attitudes of the agent who performs it. According to Dray, a motive is in one way more like a cause than Ryle realised. It is not a ghostly event, but neither is it a law-like conditional. A motive does not tell us how the agent generally behaves, but rather explains an action by identifying the agent's reason for doing it. Dray is willing to grant determinists that a reason may be a special kind of cause. He argues against Ryle that reasons, like causes, have explanatory power, while conditional generalisations do not. But a reason for him, is not a cause in the sense of an antecedent event, and it is not logically linked to its 'effect' by covering laws.

A. I. Melden's analysis of the difference between reasons and causes is both more radical and more consistent than that of either Ryle or Dray. Melden maintains that the rational explanation of an action is incompatible with any causal explanation. To ask, 'Why did X perform that action?' is to imply that X's behaviour can be described only in purposive language. Conversely, to ask for the cause of X's behaviour is to imply that X's behaviour was involuntary and that purposive explanation is inappropriate. 'Why did X raise his arm?' asks for the reason. 'What made X's arm rise?' asks for the cause.

Melden offers two main reasons why a rational or purposive explanation is incompatible with a causal explanation: (1) Because the motive for a voluntary action logically presupposes the action it motivates, while a cause must be identifiable independently of its effect. (2) Because the motive of an action is part of the way in which we identify the action. 'As motive it... tells us what the person was doing.' For example, raising one's arm in order to signal is a different action from raising one's arm in order to restore circulation.

Melden's arguments are not entirely satisfactory, however, because his first reason is inconsistent with his second reason. If the motive helps to define the action, then it cannot presuppose the action, or else our method of identifying both would be viciously circular. For we must then know the motive in order to identify the action and also know the action in order to identify the motive.

I think Melden's first argument is sound, and that his second argument is wrong. But the proof that his second argument is wrong will provide us with still another reason for agreeing with Melden's thesis that a motive cannot be a cause. The proof I have in mind starts from Melden's first claim, that a motive can only be identified in terms of the action it motivates. It follows from this that it makes no sense to speak of a person's motive before he has performed the action. A motive can only be known ex post facto, which may explain why psychologists have so much trouble predicting what their patients will do. This consequence of Melden's first argument is independently supported by the facts of language. It would be absurd for a detective to say that Smith's motive for killing Brown was jealousy, if in fact Smith has not killed Brown. (It may be objected that we sometimes do say such things as 'Several people had a motive for the crime', so that the motive seems to exist independently of the action. But here I think it is clear that 'several people had a motive' means the same as 'several people had a possible motive' — in other words, a strong, but not decisive reason. We occasionally make the word 'motive' do the same job as the word 'reason', but in such cases it is not doing its own distinctive job of signifying a motivating reason.)

Melden is therefore right in maintaining that a motive, in the strict sense of the term, must be identified in terms of the action it has motivated. But if he is right in this, then it must be inappropriate to identify an action in terms of its motive. And there is independent evidence that Melden's second argument is unsound. A man's action can have
Persons

all sorts of motives; it cannot therefore be uniquely identified by any one of them. Melden argues that a person’s motive for raising his arm—say, in order to signal—tells us what he was doing, namely, that he was signalling. But I think that Melden is here confusing what one is trying to accomplish with what one is actually doing. Usually these two things coincide and so it is not necessary to distinguish them. But sometimes one tries to signal and fails, in which case, one is not signalling, but merely raising one’s arm ineffectually. The point is that we can describe actions in two different ways. We can identify an action by what is actually accomplished, whether the result was intended or not, and we can identify it as a successful or unsuccessful attempt to accomplish some purpose. We can say that a hunter shot his friend by mistake, or we can say that he tried unsuccessfully to shoot a bear with consequences disastrous for his hunting companion. Now these two modes of identification are not on the same level. The first mode is simpler and logically prior to the second. As Melden himself points out, following Ryle, the concepts of successful and unsuccessful effort are logically dependent on the simpler concept of purposive action in paradigmatic situations (such as a healthy man raising his arm) in which the question of success or failure does not arise. You could never teach a child the meaning of failure (or success) unless he first understands what is involved in just doing something. As Ryle has put it, trying, succeeding and failing are ‘second order performances’, logically parasitic on the first order performance of just plain doing.

Setting aside the individual variations, for the moment, and considering the Dray–Peters–Melden theory of psychological explanation as a general insistence on distinguishing reasons from causes, this view suffers from one apparently serious defect. It seems to lead to the ethically unacceptable principle that every action is justifiable; tout comprendre, c’est tout justifier. The causal model of psychological explanation, which the Dray–Peters–Melden view is intended to supplant, has always suffered from a similar defect. The causal model makes every action excusable in that, given the antecedent conditions of an action, it is assumed that it is not within the agent’s power to refrain from the action. From this standpoint human action is not subject to moral evaluation because it is never rational. But if we follow Dray’s analysis of rational explanation, human action, when fully understood, is always rational and therefore always ‘the thing to have done’. For if, as Dray claims, the function of explanation by reasons is to make the agent’s conduct rationally intelligible and if ‘intelligibility’ means, as Dray seems to hold, that we can imagine ourselves doing the same thing for the same reasons, then we can never say that the agent was wrong to do what he did. ‘All reasons’, says Dray, ‘must be good reasons... in the sense that if the situation had been as the agent envisaged it... then what was done would have been the thing to have done.’ This assertion of Dray looks like a reductio ad absurdum of his theory of psychological explanation. Only the most tender-hearted social worker would agree that understanding a person’s motives requires us to approve of what he does.

Melden, too, seems to equate intelligibility with reasonableness or justifiability. An explanation of an action in terms of its motive or intention, he says, ‘reveals an order or pattern in the proceedings’. ‘To say of... a person that he is mad is to write off (his) action as unintelligible.’ Psychotic behaviour then is not intelligible behaviour and cannot be explained in terms of motives or reasons. Conversely, it would seem to follow that all actions that can be so explained are the reasonable thing to do under the circumstances as envisaged by the agent.

Richard Peters tries harder than Dray and Melden to make a clear enough distinction between justifying an action by good reasons and explaining an action by its motive, so as to avoid the absurd consequences that all psychological explanations are justifications. But Peters’s distinction is not sharp enough to do the job. Peters defines a motive as that type of reason which is called for when conduct is: (1) unconventional or non-rule-following; (2) directed toward a goal, and (3) the actual or operative reason why the agent acts as he does. Unfortunately Peters fails to explain just how we can tell which of a number of possible reasons for a person’s action is the operative one that satisfies condition (3). The trouble is that Peters, like Dray, Melden, Ancombe, P. Foot, R. Taylor and many others, fails to appreciate the fundamental correctness of Ryle’s claim that motives are bound up with character. Peters consequently rules out the only empirical means for judging whether an alleged reason is an operative one or not.

Suppose a hunter shoots his companion instead of the bear he claims to have aimed at. How do we know that he was really aiming at the bear and not at his friend? We must, of course, look first for the pattern of the events and actions surrounding the shooting, e.g. how far from his victim the hunter was standing, how the hunter behaved before and after the fatal shot, etc. But suppose an investigation reveals nothing suspicious about these immediate circumstances while, on the other hand, the investigating detective learns that the victim has been carrying on an affair with the hunter’s wife. Wouldn’t the detective then have to make an assessment of the suspect’s character in order to decide between an innocent and a guilty motive for his action? Peters and Melden make the mistake of considering only the cases where an agent’s intention is sufficiently obvious from the immediate context of the act to be unproblematic. But in cases where the motive is in doubt, we must widen the context of inquiry and thereby relate the action to deeper or more permanent traits of the agent.

Despite these difficulties, I think the Dray–Peters–Melden theory of
psychological explanation is substantially correct. It needs only to be
qualified so that psychological explanation can be clearly distinguished
from ethical justification. Closer attention to the difference between
real or motivating reasons and merely possible reasons will, I think,
allow us to combine the characteriological analysis of Ryle with the
empathetic approach of Dray.

Dray thinks of reasons as grounds for justification and he assimilates
motives to reasons. Ryle thinks of reasons as character traits and he
thereby assimilates reasons to motives. But reasons that are not motives
have no psychological reality. They are merely statements to which we
appeal in urging or supporting an action in advance (whether to our-
selves or to others) and in justifying or excusing an action in retrospect.
Reasons are logical appeals made by an appellant to a judge, even
when appellant and judge are the same person. As such, they have no
empirical relation to observed behaviour and thus no explanatory
function. To explain an action, rather than to support it or justify it,
a reason must, as Peters notes, be the agent’s real reason; it must
actually motivate the action. It is this difference between mere reasons
and real reasons or motives that tempts us to accept a causal mode of
psychological explanation. We tend to assume that the sense of ‘real’
here involved is that of spatio-temporal occurrence, as if a real reason
were an event of some kind. But there is an alternative and, I think,
more illuminating way of distinguishing real reasons from merely
possible reasons. Why not consider motives to be a special type of
reasons, namely, those to which a person could (and, in his candid
moments, would) consistently appeal in attempting to justify his actions?
My real reason or motive for my action is the reason that relates it to
other actions of mine in a wider context. If I donate money to charity
only when my name is made public, then my avowed reason, ‘to help
the needy’, although it does serve to justify my action, is not my real
reason because it is inconsistent with my refusal of funds when the
names of donors are not announced. My reasons are not, as Melden
claims, patterns of my behaviour. They are whatever statements I
make in support of my action. But their reality as motives is, indeed,
verified in terms of the observable pattern of my behaviour. Once we
distinguish real reasons from justifying but merely possible reasons, it
becomes evident that not all real reasons need be good reasons, and
that psychological explanation does not entail empathetic justification.11

If we make this distinction between reasons and motives, we need
not go as far as Dray in holding that understanding a person’s action
requires us to imagine ourselves doing the same thing under the same
circumstances. After all, psychiatrists sometimes succeed in under-
standing a psychotic patient whose behaviour is irrational from any
standpoint. It places too great a strain on the imagination of the
psychiatrist to demand that he share, even vicariously, the delusions
of his patient. Surely it is sufficient for him to disclose motives for his
patient’s actions that enable him to predict future outbursts of irrational-
ity. Yet, if motives are reasons and not causes, then how can they be
grounds for prediction? I think the answer is that in discovering the
motives (real reasons) of a person one has, in effect, disclosed his
standards of value — his likes, preferences, aspirations and measures of
importance — and one may reasonably assume that he will apply the
same standards of value in the future as he did in the past. Scientific
studies of human behaviour can thus make reliable inductive inferences
from past values and actions to future values and actions without
employing explanatory theories based on a causal model.13

Thus understanding a person’s behaviour does not require us either
to see it as causally determined or to find good reasons for it; we need
only find the real reasons for it. An action may be explicable by bad
or even absurd reasons, if these happen to be characteristic of the
agent. It would take a superhuman effort of the imagination to see
how testing one’s freedom of action could be a good reason for killing a
man, although it was Lafcadio’s reason in Gide’s novel, and it is a
major factor in the motivation of juvenile delinquents. Dray’s principle
that a reason must be good to be intelligible is plausible only insofar as
it is misleadingly trivial. Any reason by which a person is motivated is
a good reason for him in the sense that he thinks it is good. But this
truisism constrains us neither to agree that it is good nor to imagine our-
selves mistakenly thinking it to be good.

To give a good reason for an action is to justify it; to give the agent’s
real reason or motive is to explain it. But it must be granted Dray that
there is a logical relation between rational explanation and justifica-
tion. The motive by which we explain a person’s action is the reason to
which he could consistently appeal in supporting, urging or justifying
actions to himself or his confidante. “In his book” his motive justifies,
but “in our book” it need only explain.

Yet even if the point be granted that what we want of a psycho-
logical explanation is the reason for the act, and not the cause, may it
still not be the case that, since every human act is an event which can
be described in physical language, as a certain set of bodily movements,
rational explanation and causal explanation, while belonging to
different modes of discourse, are yet equally applicable to the same
entities, depending on how they are described and what we want to
know about them? May not the two kinds of explanation be com-
patible and of equal value, each in its own way — psychological, when
our purpose is to evaluate the action as right or prudent or at least
excusable, or the reverse, and physical, when our purpose is to set the
bodily movements in their appropriate place in the law-like system of
nature? In order to show that the answer to this question is nega-
tive, I shall examine the arguments for the affirmative offered by a
Persons

distinguished defender of the compatibility of these two modes of
explanation, and I hope to show that his arguments are not only in-
conclusive, but lead to inescapable contradictions.

DOING, CAUSING AND CAUSING-TO-DO

Taylor’s Two Theses

Richard Taylor’s Action and Purpose attempts to drive still deeper
the logical wedge inserted by Wittgenstein and Ryle between discourse
about human actions and discourse about natural events and processes.
Taylor explores the notions of cause, power, action, will, reasons,
deliberation and purpose in an effort to show that explaining why a
person performs an action is a quite different matter from explaining
why a water-pipe bursts or why leaves turn brown in autumn. In this
endeavour he adds fuel to the controversy that has been raging in
recent years between scientific determinists and the new wave of anti-
determinists who hold, not so much that actions are uncaused, as that,
like numbers, possibilities and jokes, they are not the sort of things of
which it makes sense either to ascribe or to deny a cause, in any sense
of ‘cause’ that would give comfort to scientific determinism.

Taylor’s discussion of the language of action is elegantly, 
refreshingly non-technical and forcefully stated, although many of his
arguments, which were previously formulated by Melden, Peters, 
Kenny and others, have been challenged by a number of writers in
defence of scientific determinism, and it is a bit disappointing that
Taylor does not trouble to meet their objections, a fault which I shall
try to rectify in the next section of this chapter. The most original
aspect of Taylor’s approach is his starting-point, although this, as we
shall see in the final section, is also the most dubious aspect. Instead of
beginning, as Melden, Peters and Kenny did, with the concepts of
action, intention and purpose, and exploring the logical differences
between these concepts and those of natural science, Taylor begins
with the ‘metaphysical’ concept of causation and tries to show that the
original (and he thinks still the correct) concept of cause was that of an
animate substance, exercising his power to produce objects or changes
in his environment. Taylor argues persuasively that his notion of cause
and its cognate notion of power cannot be reduced, à la Hume, to
mere invariant succession of events. He then attacks the volitional
theory of human action, arguing, along much the same lines as Melden
and Kenny, that alleged mental causes of actions, such as volitions,
have neither explanatory value nor intrinsic intelligibility. In the
remainder of the book he argues persuasively that the appropriate type
of explanation of human actions is purposive rather than causal, and
that the former is not reducible to the latter.

Cause and Reason

Taylor’s major thesis is that the psychological concepts employed to
explain voluntary human actions, concepts like motive, intention,
desire and reason, are not causal concepts. The explanations they pro-
vide are not predictive in function, since their use presupposes that the
agent who performed action A could have refrained from doing A.
His arguments, most of which have already been stated by the pre-
decessors mentioned, but which he formulates with special sharpness
and verve are: (1) The notion of (voluntary) human action, to which
motivational concepts apply, entails that of power or ability, and this
notion entails causal indeterminism. ‘I can move my finger’... entails
... that whether or not I do move my finger is ‘up to me’ or ‘within
my power’. (2) Motivational concepts have no predictive value, because
they are ex post facto. ‘Our entire criterion for saying what he wanted
(or tried or intended or whatnot) to do, is what he in fact did.’ (p. 52)
(3) If motivational concepts referred either to mental causes or to brain
states then the behaviour they are alleged to cause would not be volun-
tary action; it would not be what the agent does but something that
happens to him, like a sneeze or a spasm (p. 73). (4) The relation
between motives and actions is logical rather than causal, since motives
can be identified only in terms of the actions they motivate.

These four arguments which, for the sake of brevity of reference, I
shall dub (1) the ‘up to me’, (2) the ex post facto, (3) the ‘uncontrollable
reflex’ and (4) the ‘logical connection’ arguments, have a strong
intrinsic plausibility because, I believe, they are on the right track, but
as Taylor states them, they are inconclusive. The ex post facto
argument is simply a corollary of the logical connection argument; the
uncontrollable reflex argument (that if my action is caused then it is
not an action but something that happens to me like a sneeze) begs
the question, since the determinist can and does reply that sneezes are
one kind of happening and voluntary actions are another; the ‘up to
me’ argument, that voluntary action entails the power to do or to
refrain, is obscure, because Taylor insists that his notion of power is
unanalysable, and if so, we have no criterion for determining when an
agent has or fails to have such power; finally, the fourth argument,
that there is a logical bond between motivating state and action is
inconclusive because the nature of this bond is not spelled out, nor is
it proved that such a bond exists. While the logical connection argu-
ment, drawn from Wittgenstein and Melden, seems to me the most
promising, it has been powerfully attacked in recent literature, and, if
it is to be made convincing, the objections raised by determinists should
be met and the nature of the logical bond should be made clear. In the
part that follows, I shall try to do just this – to defend Taylor’s logical
connection argument against Davidson, Goldberg, Gean and others,
and to make more clear just what the logical connection is, from which
it will follow that Taylor is right to maintain that reasons, intentions,
etc. are not causes of actions. In the last section of this chapter I shall criticize what I believe to be unnecessary concessions to determinism made by Taylor, to the effect that actions may be caused by events other than motivating states; concessions which seem to me to render his entire theory of action incoherent. The reason for these self-defeating concessions on Taylor's part lies, I think, in his failure to distinguish three quite different things, namely, doing, causing and causing to do. If I succeed in establishing these two theses, (1) the pro-Taylor thesis that voluntary actions are not caused by motivating states and (2) my own anti-Taylor thesis that voluntary actions are not caused by anything other than motivating states, then it will follow that voluntary actions are not caused at all, a conclusion whose importance I need not spell out.

In Defence of Taylor's First Thesis
The Melden–Kenny–Taylor thesis, that actions are not caused by reasons, intentions, motives, decisions or purposes, as supported by the logical connection argument, has been attacked by many determinists, most notably by Davidson, Kim, and Brandt, Hempel, Goldberg, Berofsky, Gean, Margolis, Kaplan, and A. C. Macintyre. Davidson and Goldberg question whether there is in fact a logical rather than a causal connection between motive and act and argue that, even if there were, it would not be of the type that excludes a causal relation. Davidson argues that the relation between a reason or motive and an act is like that between a physical disposition such as brittleness of glass and an event such as the glass breaking under an impact. The brittleness, together with a trigger event, say the impact of a stone, causally explains the shattering of glass, and similarly, a motivating state together with some stimulus event causally explains an action. Much could be said about the defects of this analogy between psychological states and physical dispositions, but space does not allow an exploration of that point here. I want only to make the general comment that Davidson's notion of causal explanation is a tenuous bridge straddling a metaphysical chasm between two totally different kinds of explanation which, for various historical and psychological reasons, most of them consisting in philosophical mistakes, have been confused by every determinist. The two kinds of explanations generally confused are: explaining what caused event E to occur, and explaining what caused agent X to perform action A. I shall postpone the exploration of this chasm until later, pausing only to remark that Richard Taylor, disdaining Davidson's unstable bridge, tries valiantly to leap across the chasm but doesn't quite make it to the other side.

Does a logical bond between motive and act exclude a causal one? Davidson, in 'Actions, Reasons and Causes' daringly says no. He considers it a false cliché to insist that causes must be logically independent of their effects, and he denies, or appears to deny, that a causal relation is always an empirical one. With eye-raising bravado, he declares:

In any case there is something very odd in the idea that causal relations are empirical rather than logical. What can this mean? Surely not that any true causal statement is empirical. For suppose A caused B to be true. Then the cause of B=A; so substituting, we have 'The cause of B caused B' which is analytic. Davidson here interprets the claim that causal relations are empirical as equivalent to the obviously absurd claim that, to use his own words, 'any true causal statement is empirical', where, by 'causal statement' he obviously means any statement involving the word 'cause'. But few who assert the first would assert the second, and no one would regard them as equivalent. The anti-determinist's point is that no statement involving the word 'cause' is analytic (what about 'A cause is a cause?'), but only that no successful attempt at a causal explanation, i.e. no statement that significantly asserts a causal relation to hold between two events is analytic. Surely Davidson would not want to hold that his sentence, 'The cause of B caused B' provides a causal explanation of B? It seems as odd to me that Davidson should wonder why other people assert that causal relations are empirical as it seems odd to him that they do assert this.

Bruce Goldberg, in an article in Analysis, follows Davidson's lead in attacking the logical connection argument, and, trying to cross the same untenable bridge between motives and causes, falls gallantly with Davidson into the metaphysical chasm.

Goldberg argues that the apparent force of the logical connection argument (i.e. the argument that motives cannot be causes because causes are contingently linked to their effects while motives are logically bound to the actions they motivate) rests on a confusion between events themselves and the particular descriptions under which we refer to them. If C is the cause of E, C can always be described in terms of E or E in terms of C, without thereby making one a logical function of the other. Thus, as Davidson had already pointed out, we can answer the question 'What happened to Jones?' by saying 'He was burned,' thus redescribing the effect, Jones' injury, in terms of the cause, the event of his contact with intense heat. 'The truth of a causal statement', says Davidson, very sensibly, 'depends on what events are described; its status as analytic or synthetic depends on how the events are described.' This is of course perfectly true, but it completely misses the point of the Melden–Taylor logical connection argument. For the crucial point about motives is that they are not only sometimes described in terms of the actions they are alleged to cause, but that this is the only way they can possibly be described and identified.
Persons

Now Goldberg, in his gloss on Davidson, notices the difference between a description of C necessarily involving reference to E and its only happening to involve E, but he argues that even such necessary involvement does not preclude causality, since from the fact that the description of the motive necessarily involves reference to the action 'It does not follow that the occurrence [of the motive, R.A.] entails the occurrence of the action. If I want to go to the theater, does it follow that I go to the theater? There are at least some occasions when we don't do what we want to do.'

Goldberg's argument sounds eminently plausible, but only because he assumes, like most determinists, and even many anti-determinists who are muddled on this issue, that the relation between motive and action, if it is to be a logical bond, must be the relation of unrestricted entailment i.e. that a motive for doing action A must entail that the agent actually does A. And of course, such a position is plain silly. But from the silliness of this conception of the logical bond between motive and act, it does not follow that there is no logical bond at all. The true bond, it seems to me, is that of contextually limited entailment between motive and act. I shall not develop the point fully here, but briefly, it goes like this: Assume that Jones wants, intends, desires, or in some sense has a motive to open the window. What does this entail about what he will do? Well, it entails that he will open the window, but it does not entail this tout court. It entails that he will open the window provided that no reason arises for his not doing so (e.g. a hurricane is not blowing outside) and provided nothing prevents him (e.g. he is not paralysed and the window isn't stuck). The provisos here constitute the contextual limitation I spoke of on the entailment between motive and act. To say 'I want to open the window, nothing prevents me and I have no reason or motive not to, not even the motive of laziness, but still I won't open the window' is senseless. What on earth could I mean by 'want'? In this contextually limited way, a motive is indeed logically connected to an action, and not just through the way that it happens to be described, and not just to the concept of the action, but to its actual performance.

I turn now to an article by W. D. Gean entitled 'Reasons and Causes' in which Gean tries to cross the Davidson bridge with equally disastrous effect. Gean too wants to show that reasons and motives can be causes. Like Davidson, he concentrates on the word 'reason' rather than motive words like 'want', 'intend', 'decide', etc. This move of Davidson and Gean produces a particularly creaky coupling for their bridge. There is at least some specious plausibility in taking motivating states like wanting and intending to be causes of actions, but to take reasons as causes requires unusual insensitivity to the nuances of the English language. For reasons are things that need not even exist, and how can something cause something else if it doesn't exist? That this room is on fire may be my reason for rushing out of it, even if the room is not on fire. But if not, how can the non-existent fact of the room being on fire cause me to rush out? Gean and Davidson might reply that both the reason and cause of my rushing out is, not the fact of the room being on fire, but my belief that it is on fire. But this is simply an additional linguistic error. Beliefs count as reasons only in very special cases. If I believe I am Napoleon, that is a good reason for me to consult a psychiatrist. For here, the fact which constitutes a reason for my action is the fact that I believe something. But in the earlier example of my rushing out of the room, it was not the fact that I believed something that constituted, for me, a reason to rush out; rather it was the content of my belief, namely, the fact that the room is on fire, that was my reason for rushing out. To put it simply, I run away from real or imaginary fires, not from my beliefs, except in the special cases where my beliefs themselves are objects of dread, such as the belief that I am Napoleon.

Before leaving all this wreckage due to Davidson's bridge and turning to observe Taylor's fatal jump, I would like to consider one very important further point made by Gean, a point that will lead me into my criticism of Richard Taylor's second thesis in the next section.

Gean's additional argument that reasons are also causes is based on an appeal to ordinary discourse. It is, he points out, a noteworthy fact that statements explaining action in terms of reasons are translatable into statements explaining them in terms of causes. For example, the explanation, 'His reason for leaving the party early was that he had a later appointment' could be restated as 'Having a later appointment caused him to leave the party early'. Similarly, 'The reason for the ambassador's protest was the President's remarks' can be translated without loss into 'The President's remarks caused the ambassador to protest'.

Now this argument would be convincing if we were to assume, as Gean does, that the difference between A causing B and A causing B to do C is so insignificant as to be safely ignored. But the difference between causing and causing to do happens to be a deceptive looking fissure that, on further exploration, widens into a metaphysical gulf precisely as vast as that between reasons and causes. To see this, we need only look for the disanalogs between reason language and causal language, after having granted Gean the analogy on which he rests his case. With what other types of locutions does 'causing to do' (as distinguished from merely causing) freely mingle? Does not 'causing to do' belong to a family of locutions such as 'bribing', 'inciting', 'persuading', 'convincing', 'coercing', and do not all these locutions indicate different ways of doing the same general thing, namely providing an agent with a reason for performing a certain action? Now contrast this behaviour of 'causing to do' with the simpler
Persons

expression, just plain ‘causing’, as when we say that heat caused an
explosion, or the failure of the brake caused the accident. Could we
replace the sentence, ‘The sudden heat caused the explosion’ with any
substitute of the form ‘The sudden heat persuaded, incited, bribed or
coerced the explosion’? Clearly, plain causing goes with producing,
making, bringing about, and other such locutions, while ‘causing to do’
belongs to quite a different family, namely, the reason-offering family.

Critique of Taylor’s Second Thesis

After having defended Taylor’s first and, I think, true thesis, that
motives are not causes, I turn now to his second and, I think, false
thesis, that actions may have causes other than motives. But here we
must distinguish between three different contentions, the first of which
is rightly rejected by Taylor, while the second and third are wrongly
defended by him:

1. Actions are physical events caused by antecedent physical events.
2. Actions are caused or brought about by persons.
3. The agents who perform actions may be, and perhaps are always,
caused to perform them.24

Contestation 1 above has been adequately criticised by Melden,
Richard Taylor and Charles Taylor.25 The argument of Melden and
Richard Taylor is that, if actions were explainable as necessary results
of antecedent events other than the agent’s intentions or purposes, then
the agent would have to perform them whether he wanted to or not;
thus all actions would turn out to be non-voluntary. The agent would
be a ‘helpless victim’ of the forces compelling him to act, as he clearly
is in the case of a muscular spasm or a sneeze. Charles Taylor’s argu-
ment is more theoretical and, I think, more compelling: If all actions
could be explained by antecedent physical events, then there would be
no need for psychological explanation in terms of intentions, motives,
reasons, etc. Thus all genuinely psychological concepts would be
empirically vacuous. He does not deny the possibility that behavioural
sciences of the future may come up with adequate causal explanations
of all human actions in terms of conditioning processes and neural
events, but merely points out that the weight of the available empirical
evidence is overwhelmingly to the contrary, so that the thesis of
naturalistic determinism is a faith in defiance of the known facts.

Contestation 2 is an interesting ploy proposed by Richard Taylor,
who dissents the ancient conception of a person as a ‘first cause’ of his
actions. Now there would be no strong objection to this way of putting
things (which has at least the merit of reminding us that actions are not
caused by antecedent events) were it not for its misleading suggestion
(contrary to the main thesis of Taylor’s book) that actions are explain-

able in the same sort of way as physical events. It appeals most strongly
to those who are wedded to the a priori principle that everything has
some kind of cause and who reason that, if an action is not caused by
an event or process, then it must be caused by an agent. But this account
is a misleading attempt to have the best of two incompatible worlds.
For what is really gained in the way of understanding by saying that a
person causes his actions rather than just saying that he performs them
or does them? Are we not forcing language to fit a theory of causality?
Ordinarily, when we attribute causal agency to a person, we refer to
some event that he brought about by means of his actions, as when we
say ‘Smith caused the accident’. Smith’s action, in such case, consists
precisely in his causing of the accident, so that, if we were to say that
he caused his action, we would be asserting that he was the cause of
causings the accident, thus launching into an infinite regress.

The trouble with Taylor’s account of agency is that he is not clear
enough on just what he regards the agent as the cause of. The second
term of his causal relation oscillates between (a) the action, (b) the
bodily motion involved in the action, and (c) the event or state of
affairs brought about by means of the action.26 But these are very
different matters. It is one thing to assert that an agent causes his own
action, quite another to assert that he causes the bodily movement in
which his action consists, and still another to say that he causes the
state of affairs resulting from his action. The first two claims are, I
think, incoherent, while only the third makes clear sense. For consider:

(a) The claim that a person causes his action; what more does it say
than that he simply does it? If it says anything more, it suggests that
the agent makes his action happen, that he brings it about. But can an
action be made to happen or be brought about? I can make it happen
that a light goes on by raising my hand and pressing a switch, but can
I ‘make it happen’ that I raise my hand? If so, how? Where we speak
of making things happen we take responsibility for being able to
describe the means or manner by which we get them done. Yet, as
Taylor himself observes, we cannot describe how we perform what he
calls ‘simple acts’. The question ‘How?’ makes no sense. We just do
them. But if so, then it is very misleading to insist that we cause them.
Moreover, if we were really to consider an agent as the cause of his
acts, we would, as Taylor himself argues when he criticises the
‘volitional theory of action’ (yet somehow forgets when he pushes his
own agent-as-cause theory), commit ourselves to the assumption that
the agent and his act are separately identifiable entities. Yet a moment’s
reflection is enough to see that this violates the logical grammar (shall
we say ‘ontology’) of all inter-translation natural languages. We
cannot fully identify a particular action without identifying the person
who performed it. My action of raising my hand is not the same as
your (or even my) hand.
(b) Can the agent be the cause of the bodily movement involved in or constituting his action? If I raise my hand, am I the cause, not of raising my hand, but of the event of my hand rising? Taylor, following Wittgenstein, maintains that there is nothing left if we subtract my hand going up from raising my hand. But then, if it is senseless to say that I cause my raising my hand, it must be just as senseless to say that I cause my hand to rise. For these are not different events, but the same event under different descriptions. Moreover, if I am said to cause or bring about my hand's going up, as we ordinarily use this location, it would imply that I do something simpler, M, which results in the movement of my hand. If my right hand were paralysed, I might grasp it with my left and force it to go up, in which case I would indeed have caused my right hand to go up, and I could easily explain how I brought this about. But in the case of a 'simple act', i.e. an act that is not analysable into component acts, no such explanation is possible. Thus, if my act is a simple one, it makes no sense to say that I cause it, while if it is complex, then we can distinguish the simple act M, which I just do, from the later event or state of affairs, E, which I bring about by means of M. In such case, I can indeed be said to have caused E, but it can as well be said that my action M was the cause of E. Thus Taylor is quite mistaken in thinking that the notion of a substantial agent as cause of his bodily movements is essential to action discourse, for any reference to the agent as cause of E can be replaced by reference to his doing M as the cause of E, and it is once again obvious that doing is not the same sort of thing as causing.

(c) What, finally, of the agent as the cause of the state of affairs E brought about by means of his simple act M, or (more often) by a series of simple acts Ms into which the gross notion of an action such as robbing a bank is usually analysable?

This possibility has already been sufficiently explored in passing, and we have found it to be the only intelligible one of the three alternatives between which Taylor vacillates. It is, then, perfectly sound to say that I am the cause of whatever I accomplish or bring about. But it is just as sound to say that my actions Ms are the cause, so that there is really no need to revive the archaic notion of a substantial 'first cause'.

Contention 3: Having considered and found incoherent the contentions 1 (that actions are physical events caused by antecedent physical events) and 2 (that actions are caused by agents) let us turn to contention 3, the interesting claim made by Taylor that an agent may be caused by some event to perform an action and thus, in Taylor's language, may be caused to cause an action. It is primarily on this ground that Taylor argues that the differences he explores between action discourse and discourse about natural events do not disprove determinism, since the fact that an agent is a first cause of action does not preclude that he may be caused to act as he does, thus forming part of a deterministic chain. This position seems to me hopelessly inconsistent.

At various points in his book, Taylor offers four examples of situations compatible both with his analysis of action language and with scientific determinism. It will be useful to examine each of these examples separately:

(a) The case where an agent is moved by unconscious psychic forces, as when one asks a guest to leave, giving the reason that he is spoiling the party, but really motivated by a deep resentment, formed in early childhood, toward some relative whom the guest resembles.28

(b) The case where an agent is caused to perform an action (such as moving his foot) by post-hypnotic suggestion.29

(c) Being caused by some frightening stimulus to act in an excessive or irrational way, as when one grasps one's seat tightly on a ski lift when it suddenly accelerates.30

(d) Being coerced into acting contrary to one's normal desires as when one is 'driven to his action by threats'.31

(a) Are unconscious desires, phobias, etc. natural causes, or are they more like reasons and purposes which the agent happens to conceal (more or less successfully) from himself and from others? A. C. Macintyre has argued persuasively for the latter interpretation in his book The Unconscious. If he is correct, then our case (a) is not a genuine case of causality along with action done for a reason, but only an especially complicated case of the latter. On the other hand, if Macintyre is wrong and the orthodox Freudian account of the unconscious in terms of 'psychodynamic forces' is literally true, then the agent's explanation of his action as done for a reason (to prevent the guest from spoiling the party) is either insincere or self-deceptive. In fact, he was, in Taylor's own language, 'a helpless victim' of forces beyond his control. Philosophers and psychoanalysts are still divided as to which of these two interpretations is correct, but only Richard Taylor seems willing to accept both. Moreover, on the second interpretation, we can then ask, what is it that the psychodynamic forces of the agent's unconscious can be said to cause; do they cause the agent's bodily movements, or do they cause his actions? If the former, then, on Taylor's own analysis, the agent's movements are not actions but happenings, like sneezes and spasms. If the latter, then the actions can be causally explained without reference to the agent's purposes and intentions, which contradicts Taylor's major thesis that non-causal intentions and purposes are essential to the explanation of actions.

(b) Actions due to post-hypnotic suggestions raise a similar problem of dual interpretation. Some experts on hypnosis claim that we cannot be made to do by hypnosis anything that we really do not want to do. If they are right, then hypnotic suggestion does not cause us to act, but only provides us with further inducements or reasons (such as the
encouragement of a father-figure) for acting as suggested. If they are wrong, then hypnotic acts are in fact determined by psychic forces but, in that case, they occur independently of our intentions and outside of our control and thus, on Taylor's own account, they are not actions at all. The apparent plausibility of Taylor's example of hypnosis, like his example of unconsciously motivated behaviour, is due to the case with which he and we shift inadvertently between the causal and the purposeful interpretations of abnormal behaviour.

(c) Grasping one's seat on a ski lift because of a frightening sudden acceleration is a more challenging case. Taylor argues persuasively that such a response is genuinely the agent's action rather than mere happening, yet the stimulus appears to be a cause that triggers the response as a lighted fuse triggers an explosion. But here again, the plausibility of the example is due to a concealed ambiguity in the description. If I am challenged not to grasp my seat, or perhaps offered a desirable reward for refraining, and I do refrain, then clearly the action is at least semi-voluntary and it is not a necessary consequence of the frightening stimulus. If, on the other hand, no inducement can succeed in inhibiting my fear response, then, although perhaps it is too complex and skillful a response to classify it as a reflex, it is at least very much like a reflex in that it is not within my control and thus not my action in the agency sense of "my", thus not an action at all according to Taylor's analysis of 'action'. We often call such responses 'involuntary actions'. It is not to me a matter of great concern whether we call them actions or just responses.

(d) The case of coerced action is the most interesting. Can coercion such as threat of bodily or financial injury literally be considered the cause of an action? Or is it rather a strong exciting or justifying reason for doing what one is commanded to do? I believe the latter account to be correct, for three reasons: (i) Coercion-action is always mediated by the agent's beliefs and evaluations; whether or not I bow to a threat depends on my understanding the threat and believing that it will be carried out and also on my valuing my life or my fortune more than what I will lose by knuckling under. (ii) There are not now, and there is no ground for believing there ever will be, known laws that link specific coercions with specific actions. Even in the most extreme cases of bodily torture or threat to destroy the victim's family, sufficient counter-examples have already been recorded in history against the cliché that such threats are irresistible. (iii) What kinds of things could count as coercive 'causes'? Are they mere physical stimuli such as the shape of a gun or a thunderous shout? Clearly not so, since we must interpret these events as threatening before we respond to them as such. If I believe that the gun pointed at me is only a toy, I will not be coerced to open the safe even if the gun really is loaded. But then, if coercions are what we believe to be the case, rather than what in fact is the case, either they are the very 'mental occurrences' which Taylor for very sound reasons repudiates as a useless and inconsistent hypothesis, or they are not occurrences at all but simply good or bad reasons for the actions they explain.

I believe this last account to be the true one, but in any case it seems clear that coercion is not causation, except in the same metaphorical sense in which we might speak of ethical causality, as when we say: 'My sense of duty caused me to do it; I had no choice.' And surely Taylor would not be tempted to take this façon de parler at face value.

In general if, as Taylor himself contends, actions are not caused by antecedent events, but only by agents, then what can be his meaning in claiming that an antecedent event can cause an agent to perform an action? Is the entity, agent-to-perform-the-action (or, in better grammar, the -agent's-performance-of-the-action) the second term of a causal relation of which the first term is an antecedent event, such as the threat of death? Presumably Taylor is claiming just this, but the immediate difficulty faced by such a position is that there is none but a superficial grammatical difference between an action and the agent's performance of it; the action just is the agent's performance. Flicking a light switch is not an individual action but a type of action until it is individuated with respect to place, time and agency. Thus if an antecedent event of any kind, mental or physical, purposeful or mechanical, could cause an agent's-performance-of-an-action, then it could be said to cause the action, tout court. But this is inconsistent with Taylor's contention that actions are caused only by agents. Now we have already seen that it is either misleadingly vacuous or downright false to say that agents or their intentions are the causes of actions. If these alternatives are exhaustive, as it seems to me they are, then it clearly follows that actions are not caused by anything.
4 Motivation

‘Love is not a feeling. Love is put to the test, pain not. One does not say: “That was not true pain, or it would not have gone off so quickly.”’ – Wittgenstein, Zettel

If psychological description and explanation are rational, rather than causal, terminating in reasons and avowals, rather than in initial conditions and laws, then what is the relation between reasons for actions and the family of motivational concepts whose central members are ‘want’, ‘desire’, and ‘emotion’? I shall argue that motivational concepts, which are crucial to psychological explanation, are not themselves reasons for actions, but that their meanings essentially involve reasons; and then I shall try to sketch out the remaining components of their meanings so as to explain why such concepts seem to, yet in reality do not, refer either to introspectable private mental states, or to observable physiological states, nor even to Rylean dispositions, although they do involve physical states and behavioural dispositions in an interestingly complex and heretofore misunderstood way.

Let us begin at the periphery, and work our way toward the centre. Before dealing directly with the ‘internal states’ of wanting, desiring and emotion, let us first consider some derivative uses of ‘want’, etc. in explaining human actions, namely, first-person locutions of the form, ‘Because I want to A’, where the values of the variable ‘A’ are actions.

REASONS AND REASON TERMINATORS

My claim, in support of Peters, Dray and Melden, that psychological explanation is non-causal and that it is logically bound up with normative judgment or justification, may seem least plausible when one considers ‘explanations’ of the form, ‘Because I wanted to do it’ (or liked to, had a desire to, etc.). Wanting, liking and enjoying are not reasons as I have defined ‘reason’; they are not statements that one can appeal to in supporting, recommending or justifying actions. On the contrary, they are thought to motivate action in defiance of moral or even self-serving normative judgment. Thus they play no role in justification yet they do seem to perform some kind of explanatory function. Since they are not rationally explanatory, are they then causally explanatory? I think not. They are neither reasons nor causes but they are, I think, logically linked to reasons.

Prior to this century, when psychology was still a branch of speculative philosophy, psychologists and philosophers argued endlessly as to whether all actions are motivated by pleasure, desire or need. Freud was sufficiently steeped in this kind of causal psychology to search for ‘unconscious’ desires, needs and pleasures when no such ‘motives’ were obvious either to the agent or to the close observer. Ryle attempted to put this way of thinking to rest once and for all by arguing that the concepts of pleasure, want, enjoyment and the like are not names of internal events and therefore not the causal springs of action. They are, for Ryle, dispositional concepts that license predictions about a person’s behaviour. To say that one enjoys fishing, or that he wants very much to fish is, for Ryle, to predict that he will frequently go to some trouble to fish, that he will fish with energy and concentration, and that he will be restless and distracted when prevented from fishing.

Kurt Baier, in *The Moral Point of View*, classifies likes, wants, desires and pleasures as ‘self-regarding reasons’ for action, on the ground that saying ‘Because I like to’ (or want to, etc.) is a possible reply to the question, ‘Why are you doing that?’ I think that Baier is mistaken. As Melden has observed, we often respond to the question ‘Why did you do that?’ by saying, ‘No reason. I just wanted to’ (or like to). The appropriateness of the introductory phrase, ‘No reason’ cannot be disregarded. After all, I do not need a reason for doing what I like to do or want to do. If anything, I need strong reasons to refrain from such activities. But if wants, desires, likes, etc. are neither causes nor reasons, then what is the explanatory role of the locutions, ‘Because I want to’, ‘Because I like to’, etc.?

Suppose that, at a dramatic moment in a bridge game, Mr Jones’s foot is observed to come into violent contact with Mrs Jones’s shin. We ask Mr Jones, ‘Why on earth did you kick your wife?’ and he replies, ‘Because she trumped my ace’. He thus explains his action by citing a reason which, in his system of values, justifies it: If Mr Jones frequently behaves in this way there seems little doubt that he has adequately explained himself; the reason he gave is, in fact, the motive for his action. But we need not consider it a good reason.

Now let us suppose that instead of giving us a reason, Mr Jones replies: ‘No reason. I just wanted to (or like to) kick her.’ What would be the point of this reply, and what explanatory value would it have?

On the causal theory of psychological explanation, Mr Jones would have explained his action by identifying the psychic stimulus to which his action was a necessary response. But if this were so, then we would be obliged to admit that Mr Jones could not have refrained from kicking his wife, since he was the helpless victim of an irresistible psychic force, just as much as if he had said: ‘I didn’t kick her. It just
50

Persons

happened that my foot twitched violently.' Yet clearly, this is wrong. The right attitude to take toward Mr Jones's reply is to condemn him as a vicious sadist.

Would it then be correct to say with Ryle that Mr Jones has given us a law-like description of his characteristic behaviour and thereby provided us with a licence to predict that, under similar conditions, he will kick his wife again, and that he will do so with enthusiasm and with undivided attention? I think not, for two reasons. First, while we might infer all this information from Mr Jones's reply, he did not actually tell us all this. All he said was 'Because I wanted to'. Secondly, the Rylean expansion, like the causal interpretation, fails to reveal why Mr Jones's reply is morally objectionable. If Mr Jones has merely given us a true explanation of his conduct, then why do we feel that Mr Jones, in replying as he did, has added insult to injury? The trouble with the Rylean interpretation is that it parallels the causal one, and merely substitutes a physical stimulus for a psychic stimulus, while transferring the psychic aspect to the law-like relation between stimulus and response. Like the paramechanical view that he opposes, Ryle's view too leaves no place for moral responsibility, since the action remains a necessary effect of causal antecedents.

Baier's view, that 'Because I wanted to' states a selfish reason for the action, does a bit more justice to the ethical significance of Mr Jones's reply, but it is still not satisfactory. For it fails to explain why the reply is so remarkably offensive. Baier's interpretation would place 'Because I wanted to' on the same ethical level as 'Because she tramped my ace', in classifying both responses as bad reasons. But the reasons of fanatical bridge players, bad as they are, are surely less repugnant than the defiance of a sadist. 'Because I wanted to' is objectionable, not only because it fails to offer a good reason for Mr Jones's act, but more than that, because it fails to offer any reason, and it implies that no reason need be given. Mr Jones was not pleading his case for acquittal. He was denying the jurisdiction of the court.

Saying that one likes, enjoys, wants or desires to do something gives neither a cause nor a reason for the deed, but it is intimately related to reasons. Nowell-Smith and A. I. Melden have suggested that these locutions deny the need for reasons. Their function is not to explain any action, but to indicate that no further explanation is in order because the process of rational deliberation (giving B as a reason for A, C as a reason for B, etc.) terminates at this point. We might call this the limiting or 'null' case of rational explanation. 'I just wanted to' informs us that it is not our business to demand reasons, and it is sometimes the appropriate response to make. In the absence of reasons against, a man's activities are his private affair.

'Because I want to' (or like to, desire to, etc.) may then be said to function as reason-terminating locutions. They indicate, in various ways, that there is no point to further discussion, and they are appropriate responses when there really is no point; otherwise they are extremely offensive. In a rather similar way, 'Do as you like' and 'Do whatever you want' indicate that the speaker has no reasons to offer for or against any course of action favoured by the agent. 'Because I want to' is the agent's way of terminating rational deliberation. 'Do as you want' is the spectator's way.

Motivation

REASON TERMINATORS AND EXPLANATION

But reason-terminating locutions are not all alike. An adequate account of them must do justice to two features that have so far been left out of consideration: (1) The fact that some of these expressions have more explanatory power and informative value than others and (2) the striking difference of tone between say, 'Because I like to do it' and 'Because I had an overwhelming desire to do it'.

1. I began this discussion with Ryle's dispositional analysis of wanting, liking, enjoyment and the like, and I criticised it for not accounting for the reason-terminating function of these concepts. But I should like now to make use of Ryle's insights by qualifying what I have said so far. I have considered only the common features of reason-terminating locutions. But there are conspicuous differences. Some are more informative and therefore more satisfactory explanations than others. Suppose that you and I are sitting in my living-room. Suddenly I jump up and hurl my chair through the nearest window pane. You exclaim 'Why on earth did you do that?' Now suppose I reply, 'Because I wanted to'. You would be perfectly entitled to protest, 'That's no explanation!' 'Because I wanted to' is just about the most uninformative and unsatisfactory reply I could make to your request for an explanation. It tells you nothing except that what I did is none of your business. (It does not even inform you that my action was intentional, since you already assume this when you demand a purposive explanation.)

Now suppose that I say, instead, 'Because I had an intense desire (or need or craving) to do it'. You would then learn something about me, to wit, that I am a bit mad; something is going on inside me that may explode on other occasions besides this one.

If my reply were 'Because I like to throw chairs through windows' or 'Because I enjoy doing this', you would also come to understand that I am mad, and more than just a bit. For I would be informing you that, as Ryle would put it, this is the sort of impractical conduct I am likely to engage in whenever circumstances permit. Thus 'Because I like to' and 'Because I have a desire to' do more than terminate rational discussion. They also relate the action in question to one's
Persons

The general pattern of behaviour and in this respect they have explanatory value. 'Explanation' in the widest sense means showing how something unexpected might, with more information, have been expected. Reason-terminating locutions can play an explanatory role, although they provide neither causes nor reasons for action, by revealing features of one's character that make the action more predictable. Kyle's dispositional account of wanting, liking and enjoyment was therefore on the right track, although it did not go far enough because it ignored their reason-terminating role, and also because 'I like', 'I desire', etc. do not assert this general information about one's character, but rather, in various ways which I shall presently consider, they suggest or symptomatisate such information.

2. There is a conspicuous difference in tone between the expressions 'Because I had an intense desire (or craving) to do it', and 'Because I like to do it (or enjoy doing it)'. The first type of locution has a note of desperate urgency, it sounds grim and pathetic, while the second type sounds gay, frivolous and complacent. When both types are considered as reason terminators symptomatic of the speaker's character and personality, this difference in tone is easily explained. Such locutions terminate rational discussion for various motives, and, as we have seen, motives (as distinct from possible reasons) are rooted in character. These locutions do not state one's motive for an act, but they are motivated by and thus symptomatic of certain attitudes. Intense desire, need and want are alleged when one feels guilty about his action (hence the pathos), while liking, enjoyment and pleasure are alleged when one feels innocent, lighthearted or defiant. 'Because I had an intense desire (need, or craving) to do it' suggests to us that the speaker knows there are strong reasons, either moral or practical, against his action, but he was unable to control himself, not in the way he cannot control the weather, but in the way that one cannot help scratching an itch or drinking heavily. We all have our characteristic moments of irrationality. Provided that they are not too frequent and not too harmful to others, we hope to be forgiven for them. 'Because I wanted (needed or desired) passionately to do it' terminates rational argument in something like the way a defendant terminates a trial, by pleading guilty and throwing himself on the mercy of the court. Hence their tone of desperate pleading.

'Motivation

on the golf course. 'I like' and 'I enjoy' inform us about the agent's values. Unlike 'Because I had an intense desire (or need or craving) to do it', which admits that what the agent did was bad or wrong, 'Because I like (or enjoy or get pleasure from) it' implies that what the speaker has done is, in his view, generally worth doing and so needs no special justifying reasons. The things we like, enjoy or find pleasant are just those things that provide us with prudential reasons for doing anything else.

Reasons, Desires and Feelings

The role of reasons in purposive explanation

We explain voluntary actions in two ways: (1) By indicating that the action is one that the agent could be expected to perform under the circumstances (e.g. 'Because I (he) enjoy(s) doing it', 'Because it is my (his) habit to do so', 'Because it was my (his) turn', 'Because Jones asked me (him) to do it', etc.). (2) By assigning to the agent a purpose, or need, or desire and/or rule that explains his action as (for him) a necessary means to a goal G, and in this way shows that the action could have been expected as well as 'holding his conduct up for assessment', as right or wrong, wise or foolish, excusable or culpable.

The motivating purpose, desire and/or reason is logically linked to the specific goal, which is to say that the goal-state G would not be a goal for the agent, were he not, on his own view at least, rationally directed toward achieving it. The performance of the action A thus explained as, from the agent's point of view, the necessary means to achieving G, although contingently related to G itself, is entailed by the premisses constituting the explanans namely, that the agent wants, needs or desires G (minor premiss) and that (from his standpoint) A is necessary to achieve G, within the contextual limitations, or ceteris paribus conditions indicated in the previous chapter. In the light of this entailment relation, it might appear that a psychological explanation, as I have schematised it, suffers from the vacuity of Molière's 'dormative virtue' pseudo-explanation. But what saves it from such analytic vacuity is the ceteris paribus element, for the claim that these conditions hold is a contingent claim.

1. The first mode of explanation, I have argued, is not a reason-giving, but a reason-terminating response to the question, 'Why did X do A?', an explanation, so to speak, of why no reason-giving explanation is in order, insofar as the agent's action is what was to be expected under the circumstances, or insofar as the action, contrary to first appearances, was not rule-violating and thus not in need of justification or excuse.

2. Means-end Purposive Explanation. The most common form of
purposive explanation of action is: ‘X did A in order to get G’ (or, as we shall later see, ‘In doing A, X was accomplishing G’). Clearly, this formula is enthymematic. It takes tacitly for granted that (a) X wanted or desired G and (b) that X believed that A was the most convenient, appropriate or, in some sense necessary means to achieve G.

We often say, X’s reason for doing A was to get G. Is the goal of an action, then, in itself the reason for the action? Surely not, if the goal object or state is described independently of the agent’s rules and values. Objects and states, no matter how desirable, are not reasons which, as we have already noted, are abstract, conceptual entities, like corporations, numbers and laws. It has already been suggested that reasons are argument-premisses, thus propositions and rules appealed to in support of a conclusion. But if so, then what do they have to do with goals, which are states, events and objects? (If it be argued, as A. Kenny does, that a goal is not a desired object, but only the state of achieving and possessing an object, consider that an artist, creating a sculpture, painting, composition or poem, aims at bringing into existence an art object, not his own state of satisfaction. Were he to aim at the latter, he need only swallow a tranquiliser.)

I suggest that a reason is (a) a logical link between an action and its goal, in the form of a rule, ‘Ceteris paribus, doing A is necessary to get G’, or (b) a logical link between rule and action, by indicating the goal. In a practical syllogism, the full argument for the performance of A would be: ‘I (you, he) want G, doing A is necessary to get G, therefore it is necessary (i.e. right, appropriate, advisable, etc.) that I (you, he) do A.’ Either the major premiss (the rule) or the minor premiss (indication of the goal aimed at) can be considered the reason(s) for A, depending on what we take for granted, in contrast with what we want to know to satisfy us as an explanation.

For example, if we ask ‘Why did Jones open the window?’ because we want to know what Jones was trying to achieve, the reason (i.e. missing link in the explanation) might be, ‘To air out the room’. For in such case, the rule at work, ‘Opening a window is the way to air out a room’, can be taken for granted. But suppose the event occurs in London during the Second World War and the room is illuminated, so that the appropriate rule here is ‘To air out a room, it is necessary to install a blackout curtain, not to open a window’. In such a context, we would want to know what, if any countervailing rule Jones was acting on (was he obeying the orders of a treacherous superior officer, was he himself acting for the enemy?) or whether Jones was simply ignorant of the prevailing rule. It would not be an adequate explanation to say, simply, ‘He opened the window to air out the room’. One does not air out a room by getting the roof blown off by a bomb.

If we see Jones sawing and nailing pieces of wood and we ask, ‘Why are you doing that?’ we usually want to know what result he is trying to achieve. Often, in such cases, we ask ‘What are you doing?’ where the description we want classifies the simple actions we observe under a complex activity that includes the goal, e.g. ‘I am making a cabinet’. Only if Jones appears to us to be following the wrong procedure, either in pursuing his goal when he ought to be, say, mowing the lawn (for which we hired him), or in employing the wrong means (e.g. using tiny pieces rather than boards, or nails rather than screws), would we seek, as the missing reason, his rule of procedure, viz. ‘Why are you doing that?’ or ‘Why are you doing it that way?’, in which case his explanation might be, in the first case, ‘Because your wife asked me’, or in the second case, ‘Because this is the traditional method of cabinet making in my family’.

In sum, giving reasons for actions is citing goals and/or means-ends rules for achieving goals. What then is their connection with desires, feelings and emotions?

The Role of Wants, Desires and Feelings
Wanting, as we saw in the previous chapter, is a catch-all motivational term that indicates that the action it modifies is voluntary and either not in need of reasons (‘Because I want to’), or a means to some desired goal (‘Because I want G and doing A is the way to get G’), or that the agent is in some determinable state requiring action, the determinate state being specifiable as desire, need, craving, longing, compulsion, temptation, etc.

While more specific than ‘want’, ‘desire’ is still both vague and ambiguous (even more so in French, where ‘désirer’ is practically synonymous with ‘vouloir’). But ‘desire’ is more suggestive of agitation, feeling and reduced control than ‘want’, although less so than ‘crave’ or ‘long for’. We speak more easily of desires as irresistible than of wants, while craving and need entail (ceteris paribus) irresistibility, and longing entails both irresistibility and unattainability of its object. Unlike wanting, which is seldom felt, we feel our desires, cravings and longings. Needs are more complex. One may feel a need that one does not really have, or not feel one (not even be aware of one) that one does really have. Needs can be subjective or objective, illusory or real, conscious or unconscious. This is also true of wants, so that one can say, on occasion, ‘You don’t really want that, you only think you do.’ Not so for desires, cravings, etc. Psychoanalytic talk of unconscious desires and wishes is poor choice of language. ‘Unconscious needs and wants’ would be better English for the message of psychoanalytic theory. The main difference between need and want is evaluative. To say that X needs O is, to some extent, to express approval of X’s having O. Not so for wanting, except in that somewhat archaic usage as synonymous with lack, as in Robert Burns’s ‘Some have meat but cannot eat, and some that can, they want it’.
Confining the discussion to desires, let us examine their relation to feelings. Feelings, I suggest, are precisely what William James mistook emotions to be, namely, awareness of bodily agitations. What kind of awareness? The thesis of Chapter 2 on avowals suggested that feelings (which, together with preferences, intentions and beliefs, are prime objects of avowals) are not entirely passive, but are interpretive and evaluative responses to a situation that includes one's own bodily changes. A throb of remorse, or of fear, or of envy or of joy, as Ryle pointed out, is just the same throb, physiologically. The difference among them consists in what they signify to the subject who feels them, that is, in whether he regards the situation he is in as good or bad, and in his reasons for so regarding it (e.g. having done something wrong, facing danger, seeing another succeed where he has failed, or achieving a desired goal, respectively). So we see that even feelings essentially involve reasons. And only when one's reasons for his feelings are foolish are his feelings irrational. It is an almost ubiquitous error of psychologists and philosophers to have relegated emotions, desires and feelings to the realm of the non-rational. The probable source of the mistake is that 'affective states' are less under one's control than are one's thoughts, assertions and actions, and therefore somewhat less subject to commendation or condemnation as wise or foolish, admirable or culpable. But as Thomas Nagel has cogently argued in his recent book, *The Possibility of Altruism*, affective states are not entirely immune to criticism, nor to one's voluntary control, since they are grounded on beliefs and values that are subject to rational attack and defence.

Now as to desires - how do they differ from feelings? For one thing, desires are intentional, they are desires for such and such, whereas feelings are, one might say, only quasi-intentional, that is, they are feelings of - not external objects, but one’s own states, such as pain, pleasure, remorse, joy, etc. We feel our own emotions and desires, but the latter, while often the objects of our feelings, are not themselves feelings, on pain of infinite regress, nor, as Kenny and T. Nagel have shown, do they entail feelings as necessary criteria for their existence. I must be aware of a desire in order to have it, but I need not feel it (although one usually feels something related to the desire, e.g. when one desires food, one usually feels hunger; when he desires power, he feels envy toward those who have it, etc.). Desire for wealth, fame, and other abstract conditions is seldom attended by sufficient neural agitation and consequent sensation to be felt. Sensual desires are, of course, always felt, or they wouldn't be sensual. As for emotions, we shall see later that they need not be felt nor (unlike desires) need we even be aware of them (although non-awareness of one's emotions is pathological).

The important difference between feelings and desires, a difference which explains the grammatical point just made, is that desires moti-
whether Freudian or existentialist, to inform us more about, while we concentrate on the more usual and paradigmatic types of emotions.

But before terminating this digression and returning to our main theme, it may be worth noting that Sartre's perceptive account of emotions, in his early monograph on that subject, was seriously marred by his reliance both on scientific psychology and on Existenz Philosophie. Like many psychologists (e.g. Duffy, op. cit.), Sartre assumed emotions to be an irrational escape from reality and, going beyond even Heidegger and Jaspers, he defined all emotions as directed toward Being, namely, as 'magical transformation of the world.' One must go back to Aristotle, Aquinas and Spinoza to be reminded of the simple fact that emotions are usually rationally grounded and useful to us, and that without emotion, life would neither be long-lived, nor worth the long-living.

But now to return to the main issue: what sort of states are emotions, and how are they related to reasons and to feelings? I have already noted, following Kenny and Bedford, that while emotion predicates usually involve dispositions toward certain feelings (e.g. 'jealous' suggests, but does not entail, that the subject will feel throbs, pangs, or sundry sensations when, say, he notices his wife flirting with another man) such feelings are not essential to emotions. We know, for example, that Stalin was jealous of Trotsky's intellectual superiority, without having the least idea of Stalin's feelings, nor indeed any assurance that Stalin had any feelings about Trotsky. What seems beyond doubt is that Stalin was not constantly experiencing throbs or pangs at the thought of Trotsky throughout the twenty-odd years of their rivalry, although he was, throughout that period, jealous enough to make frequent efforts to have Trotsky assassinated.

The connection between emotions and feelings is, I think, similar to that described by T. Nagel's account of the connection between desires and actions, namely, that both are motivated by the same reasons and thus related through a common ground. One's reason for being jealous toward his wife is that she may be unfaithful to him, and that is precisely his reason for feeling throbs, twinges or pangs of jealousy when she goes off on a journey alone, or when he observes her flirting. Whether or not our subject actually feels throbs, twinges or pangs would, I should imagine, depend on how excitable his neuro-endocrinological system happens to be. Stalin was not an excitable man; neither was he a man of intense feelings. A man who neither shows nor avows (even to himself) intense feelings does not, so far as I can understand the concept, have intense feelings. But insofar as he does have and manifests such feelings, they are associated with emotions through their common ground of reasons.

How then are emotions grounded on reasons? Again, I shall rely on T. Nagel's helpful discussion. Reasons, I have argued, are premises for practical reasoning that guides action - they are reasons for doing, not just feeling, so that one may, conceivably, act on one's reasons without experiencing either emotion or feelings, although I would imagine that such cases are pathological and rare, and that people who frequently act unemotionally are not the best friends to have - they are the kind of heartless bureaucrats who can order mass bombings of defenseless populations for 'reasons of state.' In this respect, I think Nagel's account of moral action is too strictly Kantian and erroneously discounts the role of moral emotions (e.g. Hume's sentiment of 'benevolence') that serve to limit mechanical obedience to general rules of action. For example, the soldiers of Company G at My Lai, who fired in the air or refused to fire at all, were deterred from acting on the rule of military obedience by their emotions of pity and horror at the prospect of the slaughter of innocents. Does this mean that these emotions were irrational, i.e. not grounded on reasons? Of course not, but the reason that motivated them, namely, the rule that one ought not to slaughter the defenseless, could only prevail over the rule of military obedience with the help of the emotions of pity and horror.

But now, in maintaining that emotions can motivate exceptions to rules, as well as that reasons as rules direct action and restrain emotions, have I not returned after all to the traditional Platonic-Cartesian split between reason and emotion, despite my initial claim that they are interconnected? It looks this way, but the looks are deceiving. The resolution of this apparent inconsistency may be found in the recognition of the dispositional component of the meaning of emotion predicates, a component I argued for in the section on motives in the preceding chapter. It is not enough to say that both emotions and actions are grounded on or motivated by reasons. Unlike all but the most habitual actions, emotions involve, among other things, tendencies toward certain kinds of responses. As Ryle pointed out, the frustration of a tendency produces neural agitation, thus accounting for the fact that we so often, although not always, feel our emotions. The tendency component of emotion makes it possible for the agent to refrain from acting on a reason that directs behaviour contrary to the tendency involved, and thus provides a kind of scale for weighing competing reasons.

Thus to say that X is angry at Y, or jealous of Z, or in love with W, is not merely (as Ryle too often suggests) to attribute to X inclinations toward certain open-ended ranges of behaviour and avowal, but it surely is in part to do just that. The main additional components that Ryle overlooked are the intentionality of emotions and the normative assessments they express, both of which lead us back to reasons. X cannot be angry at Y without any reason, real or fancied, provided by Y; X must at least believe that Y has wronged him. Nor can X be jealous
or in love unless the object of his emotion has, or seems to have, features that, for him, justify the kind of behaviour characteristic of these emotions.

In what way then can emotions inhibit or support reasons for actions rather than merely epiphenomenally reflecting such reasons, as Nagel seems to hold? My suggestion is this: an emotion is a tendency to weigh certain reasons above others in accordance with one’s beliefs and values and in relation to a given situation. To ascribe an emotion to oneself or to another person is, above all, to provide information about one’s own or another’s scale of values. Our emotions, which manifest the relative weight we assign to reasons for action, distinguish us as unique individuals, while our intellectual and moral judgments stem from our common human nature, i.e. from our common ability to recognize and respect reasons. There are, of course, limits to the individual variability of emotions, particularly in the case of the distinctly moral emotions such as sympathy, pity, loyalty, gratitude, etc. for these are (ideally, if not in fact) evoked in all to the same degree by the same situations. But this is, at best, true only in simple situations where only one such emotion is appropriate. More often, they come into conflict, the resolution of which then varies with the individual. Where love of one’s family motivates disloyalty to one’s political party or nation, it would be callous to denounce the victim of so agonising a conflict of values, no matter which way he resolves it. Sartre wisely declined to advise a young Frenchman who faced a choice between fighting in the Resistance and taking care of his dependent mother. Insofar as patriotism and filial love qualify as emotions, two individuals will differ in the degree to which they are moved more by the one than by the other, with respect to the particular objects of these emotions.

To sum up our findings, it is misleading to think of emotions as states of either psychic or neural agitation, or as irrational responses to disturbing situations. To attribute an emotion like anger, love or jealousy to oneself or to another is to explain his present and likely future actions in terms of the way he envisions his situation, the way he interprets his bodily agitations (if any), the goals he pursues, the relative values he places on those goals, and the rules of action that, for him, link means and ends. In brief, ascribing an emotion to a person is short-hand for an extraordinary amount of information about him, which may help explain why adequate psychological understanding is so difficult to achieve.

It is of course very tempting to visualize emotions as causal forces, as neural or psychic ‘springs of action’ that push us or pull us, impel or propel us into action, on the assumption that human actions, like natural events, are to be explained by antecedent causes. This temptation, due to our penchant for causal explanation, is further intensified by the frequent attendance of emotions by bodily agitations, and the

phenomenological fact that, when in an intensely emotional state, we usually feel impelled by some inner force that is not within our rational control.

But if we are to understand psychological discourse in a clear enough way to formulate adequately subtle criteria of verification, we must recognize this way of thinking to be metaphorical and misleading (overemphasising, as it does, the pathological cases of emotion as against the normal cases) and, like good phenomenological psychiatrists, attempt to discover a person’s reasons for his emotions and actions, rather than blocking explanations with: ‘Poor chap, he just can’t help himself, he’s driven’ (by what? an internal piston, a demon, an irresistible psychic push or a neural explosion?), a causal model that will get us nowhere in understanding ourselves and others.

H. Mullane has maintained that the unconscious reasons of a psychopathic person are not ‘his’ reasons but are somehow imposed upon him by some causal process that Mullane suggests may be ‘neurophysiological’. Such reasons are therefore ‘non-rational’ reasons. The very notion of non-rational reasons strikes one as unnecessarily paradoxical. Mullane wants to explain why compulsive behaviour is to some degree excusable without justifying it as rational, even from the standpoint of the agent’s beliefs. I have already indicated, in the discussion of Dray and Peters in the preceding chapter, how we can avoid such absurd relativism of justification without paying Mullane’s price of non-rational causal explanation. To revert to causes, whether psychic or neurological, exculpates completely, and this is a consequence that Mullane very much recognises, toward the end of his essay, and then tries desperately to escape by weakening his deterministic account of compulsive behaviour, thereby throwing in the sponge: ‘It is only the general neurotic pattern of behaviour that can be legitimately taken as caused by unconscious processes over which the agent has no control.’

A simpler way out of the difficulty Mullane rightly points to, of partially excusing neurotic behaviour while still explaining it in terms of reasons (bad reasons, yet not bad because of ignorance but bad because of deliberate self-deception (see Chapter 7)), is that ‘psychic compulsion’ is a metaphor for the natural tendency to avoid suffering. If a kleptomaniac finds relief from frustration, guilt or a sense of impotence by stealing, we pity him enough to assign him therapy rather than punishment, as we do (or should do) for a narcotic addict, without having to deny that he did, voluntarily, steal or take dope. The therapy should aim at helping the agent to find more acceptable and less self-destructive ways of satisfying his needs, thereby reshaping his values and goals into a more ‘integrated’ (i.e. rational) structure. But again, there is no necessity to drag causality back into the explanatory picture.