Aristotelian Society Series

Volume 1
COLIN MCGINN
Wittgenstein on Meaning:
An Interpretation and Evaluation

Volume 2
BARRY TAYLOR
Modes of Occurrence:
Verbs, Adverbs and Events

Volume 3
KIT FINE
Reasoning with Arbitrary Objects

Volume 4
CHRISTOPHER PEACOCKE
Thoughts:
An Essay on Content

Volume 5
DAVID E. COOPER
Metaphor

Volume 6
DAVID WIGGINS
Needs, Values, Truth:
Essays in the Philosophy of Value
Second Edition

Volume 7
JONATHAN WESTPHAL
Colour:
Some Philosophical Problems from Wittgenstein
Second Edition

Volume 8
ANTHONY SAVILE
Aesthetic Reconstructions:
The Seminal Writings of Lessing, Kant and Schiller

Volume 9
GRAEME FORBES
Languages of Possibility:
An Essay in Philosophical Logic

Volume 10
JONATHAN LOWE
Kinds of Being:
A Study of Individuation, Identity and the Logic of Sortal Terms

Volume 11
JIM HOPKINS AND ANTHONY SAVILE (Editors)
Psychoanalysis, Mind, and Art:
Perspectives on Richard Wollheim

Aristotelian Society Monographs Committee:
Martin Davies (Monographs Editor); Thomas Baldwin;
Jennifer Hornsby; Mark Sainsbury; Anthony Savile

Psychoanalysis, Mind and Art

Perspectives on Richard Wollheim

Edited by
Jim Hopkins and Anthony Savile

BLACKWELL
Oxford UK & Cambridge USA

BF 173
.P7760
1992

Copyright © Basil Blackwell Ltd 1992
First published 1992

Blackwell Publishers
108 Cowley Road
Oxford OX4 1JF
UK

238 Main Street, Suite 501
Cambridge, Massachusetts 02142
USA

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

Except in the United States of America, this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

British Library Cataloguing in Publication Data

A CIP catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Psychoanalysis, mind, and art; perspectives on Richard Wollheim/
edited by Jim Hopkins and Anthony Savile.
p. cm.—(Aristotelian Society series; v. 11)
Includes bibliographical references and index.
ISBN 0-631-17571-7 (alk. paper)
1. Psychoanalysis. 2. Art—Psychology. 3. Philosophy of mind.
4. Wollheim, Richard, 1923– . 5. Hopkins, James. 6. Savile,
Anthony. I. Series.
BF173.P776 1992
150. 19'5–dc20 91-46360 CIP

Typeset in 10 on 12 Plantin by Pure Tech Corporation,
Pondicherry, India

Printed in Great Britain by T.J. Press Ltd, Padstow, Cornwall

This book is printed on acid-free paper

Contents

Preface vii

Part I: Psychoanalysis, Values, and Politics

1 Psychoanalysis, Interpretation, and Science
Jim Hopkins 3
2 The Nature and Source of Emotion
Sebastian Gardner 35
3 Acting on Phantasy and Acting on Desire
Hanna Segal 55
4 Knowing and Valuing: Some Questions of Genealogy
Marcia Cavell 68
5 Naturalism, Psychoanalysis, and Moral Motivation
Samuel Scheffler 87
6 Three Types of Projectivism
A. W. Price 110
7 Aggression, Love, and Morality: Wollheim on Rousseau
Nicholas Dent 129
8 The Future of a Disillusion
G. A. Cohen 142
9 Character, Mind, and Politics: The Socratic Case
Amélie Oksenberg Rorty 161
10 Utopia and Fantasy: The Practicability of Plato's Ideally Just City
M. F. Burnyeat 175

Part II: Bradley and Green

11 Bradley and Moral Philosophy
Patrick Gardiner 191
12 Inscrutability of Reference, Monism, and Individuals
Hidé Ishiguro 205
13 Motions of the Mind
W. D. Hart 220

1

Psychoanalysis, Interpretation, and Science

JIM HOPKINS

Since this is a volume for Richard Wollheim, I should like to say something about working with him. We gave a series of seminars on philosophy and psychoanalysis together for a number of years while he was in London. Many of our colleagues were from the department Richard had built, and were particularly able and enthusiastic. And I had the opportunity to learn from a pioneer in our subject, whose knowledge was unrivalled, and who conveyed what he knew with a subtlety and wit which made learning a particular pleasure. It was one of the most rewarding experiences of my academic life.

One of the many things that made an impression on me was Richard's ability to use interpretive descriptions. In his presence, or reading his work, there could be no doubt that such thinking was philosophically informative. And although the illumination given by interpretive thought may derive from a particular way of seeing things, the capacity to give information in this way is not unique, and that needed to receive it, fortunately, is spread wider still. Hence, very many instances make us aware of the value of interpretation, and it seems a basic and pervasive source of understanding.

This suggests questions as to how interpretation relates to other sources of knowledge, and particularly to science; and these have often focused on the work of Freud. Psychoanalysis, as Freud envisaged it, was to be an interpretive science. But it is natural to wonder how far a discipline can be interpretive and also scientific, or objective. This is a larger topic than can be treated here, but I should like to sketch some lines of thought which contribute to an overall view.[1]

I

Let us begin with cases in which to interpret something is to understand or specify its meaning.[2] Here interpretation seems a purely hermeneutic or semantic process. The paradigm objects of such interpretation are linguistic – things like utterances, inscriptions, or texts – and the abilities we employ in it are familiar and striking, and show most clearly in our understanding of language. In particular, human beings seem naturally able to understand novel sentences in unlimited number, readily and without effort, provided these are composed of familiar words in grammatical patterns.

This evidently involves abilities which are both syntactic (concerned with relations of symbol to symbol) and semantic (concerned with relations of symbol to thing symbolized). Thus understanding words involves relating them to one another to form sentences and also relating them to things they designate. Someone who understands a noun-phrase, for example, not only knows how to combine it with verb-phrases to form sentences, but also (at least often) knows that it is supposed to stand for a certain person, place, or thing; knows how to identify such things; to communicate with people who can tell him or her about them; and so forth. And again, someone who understands a descriptive word or phrase will know that it is true of certain objects, will often be able to recognize the properties with which the description is correlated, communicate with others who can, and so forth.

Similarly, understanding sentences involves relating them to one another and also to the world. A speaker who understands a particular indicative sentence, for example, knows many relations of implication which hold between that sentence and other sentences, pairs of sentences, and so on. That is, for many sequences of sentences[3] 'S_1' to 'S_{n+1}', a speaker knows something of the form

If 'S_1', 'S_2', . . . 'S_n' are true, so is 'S_{n+1}'.

The importance of such sequences is not merely linguistic; they also serve as patterns for thinking. Thus someone who understands English will know, for example, that if both 'If cricket is a game, cricket is good' and 'Cricket is a game' are true, then so is 'Cricket is good'. More generally, a speaker will recognize that this sort of connection holds for all instances of the same pattern (here, say, the pattern: if S_1; S_2; S_1; so S_2) And such patterns will describe transitions, from sentence to sentence or thought to thought, which speakers acknowledge as cogent.

Also, a speaker who understands an indicative sentence commonly knows the circumstances which would render that sentence true. Thus a speaker of English will know that 'Snow is white' is true just if snow is white, that 'Grass is green' is true just if grass is green, that 'Dogs bark' is true just if dogs bark, and so on, without end. (This may seem so obvious as to go without saying; but this is because we are taking examples from a language we understand. We could not list truths this way for a language unknown to us, and this indicates that linguistic understanding and knowledge of such truths are connected.) So for each sentence 'S' that a speaker is able to understand, he or she will also know something of the form

'S' is true just if S.

This pattern, like the previous one, evidently has indefinitely many instances; and very many of these reflect abilities to link sentences with the world – that is, with the actual objects or situations which the sentences could be used to describe, and which would render them true or false. Thus suppose someone utters, say, 'This snow is white', in order to say of a particular chunk of snow that it is white. A person who understands the sentence will know the kind of thing that would make it true, and in consequence will know how to relate the utterance to its worldly context, and (in many cases at least) how to determine whether it is true or false. So an understander will be able to recognize the particular snow in question as verifying the utterance, and would in the absence of anything appropriate be able to judge it false. Here the white snow spoken of is something of the sort which corresponds to the sentence if it is true. And in understanding language, it seems, we are able to grasp indefinitely many potential correspondence relations of this kind, which enable us to use language as a means of communication about the world.[4]

There is clearly much more to be said about what speakers know about their words and sentences, but this may serve to start our sketch. In considering interpretation, we can usefully focus on these relations between words and things and between sentences and the events or situations which render them true. These seem central to meaning for (among others) two reasons.

First, the possession of meaning by words and sentences seems one manifestation of the broader phenomenon of representation. To specify the meaning of words and sentences seems to be to characterize

their capacity to represent how things are in the world. But to be a representation seems to be to stand in relation to something: namely, that which is represented. So it is natural to hold that, in describing relations of words to things and sentences to situations, we are elucidating meaning by relating representations to what they represent.

Secondly, stating the referents of words or the truth-conditions of sentences already seems a way of saying what they mean. Thus we could use the specifications of truth-conditions above to interpret 'Snow is white' as meaning that snow is white, 'Grass is green' as meaning that grass is green, and so on; and this would be correct. Again, if someone wants to know what a name or a descriptive word means, we can often explain this by specifying (or providing some other means of acquaintance with) the object for which the name stands, the property or relation correlated with the description, and so forth.

Our everyday interpretation of language proceeds spontaneously and without reflection. The same process, however, can be mediated (or partly replaced) by explicit hypotheses or theory, as in the case of someone who uses a dictionary or a grammar in a foreign country to help with understanding what people are saying. We can call this *theoretical* interpretation; and since it involves explicit formulations about language which can be tested in practice, we can use it to cast light on what we ordinarily do naturally. In particular, where we can interpret the same things both naturally and by means of theory, we may be able to regard the theory as specifying some of the information we bring to bear in unreflective practice.[5]

Now our ability to understand sentences on the basis of their words and the way they are put together has suggested to many that human beings must somehow grasp or otherwise embody a system of rules in accord with which this task is accomplished. These would be rules by which we relate language to the world and also, it seems, in accord with which we think. Such a vision inspired Wittgenstein's *Tractatus Logico-Philosophicus*, in which he gave an account of language and representation which was a precursor of many modern views. Wittgenstein thought that our ability to understand sentences showed that they were what he called 'pictures' of reality, by which he meant that they were representations which were both *compositional* and *referential*. And although his ideas here are not unfamiliar, they are relevant enough to our purposes to merit some spelling out.

Wittgenstein argued that in very many cases we can find what we may call systematic element – element: combination – combination correlations, as between representations and the things or reality they

represent. As examples he cited, in addition to language, alphabetic notation considered as describing sounds; hieroglyphics; musical notation; maps, blueprints, and models; and even information-bearing mechanical structures such as the groove on a gramophone record, the roll in a player piano, or the system of holes in the cards of a programmable loom.[6] (This was before the prominence of the computer metaphor.)

In each of these cases, Wittgenstein held, we can regard *both* the representations *and* the things they represent as composed of elements which are combined in certain ways. This means that the elements which figure in the representations can be correlated with elements in the represented reality in such a way that the modes of combination in the representations map on to those in the reality. Such element – element: combination – combination correlations enable representations to carry information about what they represent in a particular way.[7] And given such a system of correlations, a combination of symbolic elements can be used because it represents reality accurately; that is, because this combination of symbolic elements is mapped with a combination in reality which actually obtains.

Wittgenstein took human language to be a system of this kind, in which the combinations of symbols which we regarded as grammatical sentences had been naturally constrained (by inbuilt rules of combination and projection) to map ways in which objects, properties, and relations might be combined in the situations we encountered in reality. And his view seems to have been that we naturally *embodied* such rules – that they were 'part of the human organism' – so that we actually knew little about them apart from the connections between words and things on the one hand and sentences and situations on the other, which were the socially salient aspects of their input and output. Thus, as he says:

> Man possesses the ability to construct languages capable of expressing every sense, without having any idea how each word has meaning or what its meaning is – just as people speak without knowing how the individual sounds are produced. Everyday language is a part of the human organism and no less complicated than it.[8]

Wittgenstein thought that the philosophical clarification of thoughts should make such elements, combinations, and 'rules of projection' manifest. Although he did not think of this as a matter for empirical theory, much contemporary research can be seen in terms of this task.

To describe representations in such a system, we must be able to survey relations among representing elements themselves, and so, in the case of language, the 'rules of combination' in accord with which sentences are made up of words. As is familiar, linguists like Chomsky have tried to specify such rules fully and precisely, and have found them to be complex and abstract. Such work provides a detailed explication of this aspect of our practical mastery of language. And explications of this kind can suggest descriptions of computational mechanisms for performing such tasks, and hence hypotheses as to the working of the mind or brain.

Describing the 'rules of projection' which relate human language to the world requires specifying the referents of words and the ways these determine the truth-conditions of sentences. Despite a number of differences, touched on below, this is closely connected with another contemporary project, first advocated by Donald Davidson: that of setting out a *theory of truth*[9] for a language we wish to understand in a theoretical way.

Such a theory, again, provides a description of a language which is compositional and referential. Applied to a particular sentence of the language for which it has been devised, the axioms of a (properly constrained) theory of truth would take as input the subsentential expressions making it up, the way they are put together, and the way they are related to objects,[10] and would yield, as output, conditionals like '"Snow is white" is true if and only if snow is white', covering each sentence of the language. Such conditionals would state, for each sentence, the conditions in which it is true. We have just noted that such conditionals can be seen as describing semantic knowledge a speaker has about the sentences of his or her language, and that stating the truth-conditions of sentences is a way of interpreting them. In light of these and other considerations, Davidson has urged that such a theory could be used for the theoretical interpretation of natural language, and hence to cast light on linguistic meaning itself.

A theory of the kind we are considering regiments information about the structure of sentences and the connections between words and things in such a way that this suffices for interpretation. (Also, such a theory can be used to specify something like a manual of translation between the language of the theory and the language to which it is applied.) Hence, in so far as meaning is what interpretation reveals or what sentences and their translations have in common, it seems, once more, that there is reason to regard the information thus regimented as constitutive of meaning. This also accords with the later Wittgenstein's conception of meaning as use. Wittgenstein

stressed that the meanings of words and sentences are fixed by such public features of their use (including ostensive definitions) as an explorer in an unknown country could employ to interpret the language of a people quite strange to him. (Cf. *Investigations* §§32 and 206, and his repeated arguments that there could be no such thing as an uninterpretable language.) Davidson envisages a theory of truth being used in precisely this way; that is, for what, following Quine, he calls 'radical interpretation'. So the information regimented by such a theory as Davidson envisages is precisely that which would relate use, as the later Wittgenstein conceived it, to the more traditional sorts of specifications of meaning which Wittgenstein himself employed in his earlier work.

Compositional or combinatory semantic theories are well known, and have been studied in detail. Also, their use in theoretical interpretation is well established. They can serve to interpret codes, artificial languages, and natural language itself. They provide theories which are both rigorous and powerful, yielding an unlimited number of theorems for interpreting complex symbols on the basis of lists of axioms which are finite and precisely specified. So it seems that we can regard the process of constructing and testing such a theory – and hence that of finding and articulating meaning – as analogous to that in other explanatory enterprises, and hence, in a broad sense, as scientific.

Alan Turing stressed the analogy for a particularly clear case, arguing that 'There is a remarkably close parallel between the problems of the physicist and the cryptographer. The system on which a message is enciphered corresponds to the laws of the universe, the intercepted messages to the evidence available, and the keys for a day or a message to important constants which have to be determined.'[11] Apart from some points about the role of law to be noted below, there seems no objection to regarding theoretical interpretation in something like this way generally. In this perspective, questions as to the detail of semantic theories – for example, regarding whether a theory should include explicit reference to properties as correlated with predicates or facts or situations as correlated with sentences–are to be settled ultimately by reference to their role in science taken as a whole.

Scientific theories seem is general to be supported by their capacity to explain the data they cover, so that when we judge that a theory is confirmed, we are making an inference to the truth, or acceptability, of our best explanatory hypothesis.[12] On this account, we support a scientific theory by showing that it explains certain things well, and

better than any other; but there can be no guarantee that some further theory may not do better, and so nothing like final scientific proof. In practice we would take an interpretive theory as confirmed by success in making sense of the utterances or text to which it is applied – as in the case, say, of interpretations of coded German broadcasts. Confirmation of this kind can also be regarded as an instance of inference to the best (psychological) explanation, where what is explained is the production of the symbols which are being interpreted (in the case of Turing's group, the production of certain strings of code) and the explanatory hypothesis is that they were produced intentionally, in order to represent things as the theory specifies (produced, for example, to say *these* things, to communicate this information, these orders, and so on.).

The interpretation of human texts or utterances is also perforce that of the linguistic activities of persons. So the best interpretation of an utterance or text will represent the meaning to be assigned to it as enjoying the best possible fit with the motives – and hence with the actions, surroundings, and lives – of those who promulgated it. For this reason, we must see the application of an interpretive theory as part of an overall project of understanding both language and action. Still, within this context a particular interpretive theory can be strongly confirmed apart from detailed considerations about action, simply by its capacity to render text coherent.

This is because the interpretation of a text subjects it to the constraints of order which are essential for meaning, and these may be rigorous enough in practice to support a particular interpretation and to rule out all natural alternatives. What we are calling 'text' can be taken to consist of strings of symbols which, prior to interpretation, would seem to enjoy unlimited possibilities of combination. Under interpretation, these strings, first, must fit a grammar which restricts their combinations radically, to those of legitimate words and sentences; and secondly, these combinations must also meet the further condition that they cohere, each with all the others, as parts of an intelligible sequence. Unordered strings do not satisfy such conditions. And although here, as elsewhere, a theory may be underdetermined by the data it is used to explain, in practice a theory which makes good sense of a testing stretch of text will often lack rivals which propose seriously different interpretations.

Some of the principles at work in this can be seen clearly in simple but characteristic cases. Thus consider codes which map numbers on to letters of the alphabet. For example, we may take

1	2	3		4	5	6	7	8		9	10	11	12	13
T	h	e		q	u	i	c	k		b	r	o	w	n
14	11	15		16	5	17	18	3	19		11	20	3	10
f	o	x		j	u	m	p	e	d		o	v	e	r
1	2	3		21	22	23	24		19	11	25		22	
t	h	e		l	a	z	y		d	o	g		A	
17	11	17	3	13	1		12	3	13	1		9	24	
m	o	m	e	n	t		w	e	n	t		b	y	
9	5	1		1	2	3		19	11	25		19	6	19
b	u	t		t	h	e		d	o	g		d	i	d
13	11	1		26	1	6	10		22	1		22	21	21
n	o	t		s	t	i	r		a	t		a	l	l

Strings of numbers which can be taken as encoding words and sentences in this way are subject to familiar interpretive constraints. Each string must be so ordered as to yield a word; in addition, each string must be related to others, so that all cohere as parts of an intelligible message. Even in this short example it seems likely that such constraints make it possible to fix on a single good theory, as the reader can verify by trying alternatives. So we see that combinatory interpretive hypotheses can be quite quickly and strongly confirmed in practice, even though they deal with what might seem a daunting number of possibilities. Also, we can see that interpretive confirmation displays a number of characteristic and significant features.

First, confirmation of a particular hypothesis comes only when it is used together with a number of others, and to cover a certain critical amount of material. If we take just a little text – say, the '123' with which the above begins – very many distinct hypotheses will apparently serve equally well to render it coherent. But as more strings are taken into account and hypotheses are framed to cover these as well, alternatives tend to be eliminated, and with increasing rapidity. This suggests that in general the cogency of a good interpretive theory can be recognized only by someone who has applied enough of it, and to enough material. To someone who has not done this, a correct account may seem no better than many incorrect rivals.

This is a consequence of the nature of the theories with which we are concerned or, again, the kind of information or knowledge they embody. Any interpretive theory will, so to speak, spread itself relatively thinly over the material to be interpreted; it will require a distinct hypothesis specifying the referent, or some relation to objects, for each combining symbol, and also a separate hypothesis, or series

of hypotheses, bearing on each significant mode of combination. Plainly a theory consisting of so many hypotheses, each relating to potentially distinct data, must be applied as a whole, and to a good range of data, in order for each hypothesis to be tested in use. In this, it is worth noting, interpretive theories differ from paradigms like Newton's, which cover their data by means of just a few laws, all of which may be brought to bear at once in explaining a single event.

Once engaged, however, diverse interpretive hypotheses can work interactively. That is, claims about the referents of symbols can serve in co-ordination with claims about their significant combinations, so as to determine interpretations via various constraints of order simultaneously. Successful interpretive hypotheses can thus lock in on data co-operatively and with some rapidity, as the contribution of each is confirmed by its coherence with others. (Thus in the example above, the interpretation of '123' fixes other occurrences of those symbols, and so constrains further interpretations throughout the text; and that of the first line provides reason for fixing the reference of many numerals on the basis of just one occurrence.) So successful interpretation can yield a relatively full theory, and one which readily makes sense of new strings, with a speed which might seem surprising, given the complexity of both the data and the theory involved.

Secondly, interpretive theories illustrate the possibility that a theory can be entirely satisfactory, but yet predictive only in a delimited way. Confirmation of such a theory enables us to predict that we can use it to make sense of future strings in the same language. But even a theory which we knew to be correct and adequate in all respects would not enable us to predict future strings themselves. This too is in the nature of the case, for an interpretive theory specifies only how strings must be formed if they are to carry information interpretable by that theory. There is no restriction on the infinity of possible messages or on the production of other strings generally.

Here again, owing to the phenomena with which it deals, interpretive theory contrasts with the Newtonian paradigm. We see this also in our own grasp of language, which enables us to understand, but not to predict, what others say (and only if they are vocalizing so as to be understood, and in the right language). Also, this seems to accord with the natural idea that the function of interpretation is to enable us to make use of information provided by others. For this we do not need to predict others' behaviour generally, but rather, only to understand those aspects of it which function to pass on information. Limited as this task is, it is plainly very important.

This links with a third point. Interpretation assigns properties to strings – being generated in accord with certain rules and composed of symbols with certain referents in reality – which are highly theoretic and abstract, and which concern relations with things which may be spatially or temporally very remote from the symbols themselves. Strings will, however, show other, simpler properties of order, which an interpreter may make use of. (Indeed, number-for-letter codes are often unravelled via well-known statistics concerning the frequency of occurrence of various letters and combinations in written English.) But it would be wrong to suppose that an interpretive theory should be assessed in terms of these surface properties, as opposed to its success in yielding interpretations. For the deeper combinatory and referential order which successful interpretation reveals will serve to explain the order (or apparent disorder) which emerges at other levels as resulting from the encoding of the particular message it reveals in the particular system it specifies; but not vice versa.

This can be partly brought out in terms of the example above. Here the sequence '22 23' occurs once, encoding the 'a z' of the word 'lazy'. Since 'a' is a relatively common letter but 'z' not, '22' occurs in other sequences, but '23' does not. Indeed even a much longer message might well contain no other occurrences of 'z', or only occurrences preceded by 'i' (as in 'realize', and the like) but not by 'a'. So, plainly, the combination '22 23', or its interpretation 'a z', might be unique in a considerable stretch of text, and so anomalous by comparison with occurrences of these numerals or letters elsewhere. Still, it is clear that an interpretation which held that '22 23' was a combination with sense and which interpreted it as encoding 'a z' might be correct, and supported in the highest degree. Also, it is clear that this support would be provided in part by '22' and '23' occurring as symbols interpretable by 'a' and 'z' in other combinations. That is, the very same evidential basis, relation to which made the occurrence '22 23' an anomaly, might also render the hypothesis that it encoded 'a z' irresistible.

Since the best interpretation – and one that can be very strongly confirmed – can explain the occurrence of the sequence '22 23' on the hypothesis that it encodes 'a z', it is clear that:

1 It would be a methodological error to argue against this hypothesis on the grounds that 'a' is not usually followed by 'z', nor 'z' usually preceded by 'a', either generally or in this message. That is, it would be a mistake to allow considerations of surface symbol frequency to dominate interpretation in a case in which,

as here, the best interpretation also provides the best explanation of these observed frequencies.

2 It would also be an error to argue against this hypothesis on the grounds that it implies that the production of '22' at this point in the message was a cause of the production of '23' after it, whereas the application to this material of some non-interpretive quantitative method for determining causal connection (say a Millian or Baconian method) does not show evidence of a causal connection between the occurrence of '22' and that of '23'. For, to hold that 'lazy' is the best interpretation of the string in question is to hold that the best account we can give of the intentions (causes) which produced the text includes an intention to encode 'lazy', and hence to encode 'a' followed by 'z' and hence to write '22' followed by '23'. And this does imply that the writing, say, of '22' was a cause of the writing of '23' – in the sense in which the production of one part of a message is among the causes of that of another – despite the failure of non-interpretive methods to bring this out.

A similar point also arises for natural language. As Chomsky has stressed, many of the sentences produced by each of us in speaking our own language are novel, that is, different in one way or another from any produced on any other occasion.[13] Many uttered sentences, therefore, can be regarded as anomalies, whose character is partly owed to the information they serve to communicate. Clearly this does not make their interpretation uncertain. Rather, as with the message above, we recognize that the combinations make sense and that we understand them, and hence that the anomaly is superficial. So it would also be wrong in the case of natural language to allow surface causal or correlational considerations to dominate interpretation. (Consider, for example, the effect of the stipulation that a sentence should not be taken to be true in a particular situation unless it had been used in just that situation before, or often enough to count as a statistically significant indicator of it.)

This point also links with the nature of interpretation and communication. Combinatory strings carry a variety of information precisely by using a variety of symbols and orderings within those permitted by their compositional rules. From any but an interpretive perspective, the variety essential to the encoding of information is apt to seem anomalous, as in the case of 'a z' above or novel sentences generally. So it is essential that the presence of the kind of generative causal order involved in meaning be judged by interpretation, or the

relatively full application of a (compositional and referential) semantic theory, rather than by a method which treats of more concrete superficial features. It is a general feature of explanatory theories that they enable us to see that apparent anomalies in observation are in fact aspects of a deeper order; and in the case of interpretive theories, this is the order which sustains meaning. So methods which impose a non-interpretive filter on phenomena with meaning are liable to screen out the very information these phenomena are designed to carry.

II

We have seen that the understanding of language shows a number of distinctive methodological features, which can be related to the kind of theory which makes interpretive information explicit. I want now to argue that psychological understanding is so interwoven with that of language as to share these features. It follows, I think, that commonsense psychology and theories based on it (including Freud's) are liable to systematic mis-evaluation, through methodological judgements which are counterparts to the errors about interpretation just described. Moreover, there seems reason to hold that such mis-evaluation (or at least a bias towards it) is endemic to an established tradition in the philosophy of science.

We have reviewed the fact that people frequently know about the referents of their words and the implications and truth-conditions of their sentences. Let us register this by saying that speakers have knowledge of *semantic connections* which hold for their language, where these include both logical connections and the links of words and sentences to their worldly correspondents (or, as we may say, their *semantic values*).[14] Then we can try to bring out something of the way in which knowledge of semantic connections is bound up with psychological understanding.

As is familiar, we use language in a particular way in articulating[15] the contents of the mind. We have, first, a series of words which we take to describe motives, or mental states or acts – words like 'remember', 'expect', 'imagine', 'believe', 'desire', 'hope', 'fear', 'wish', and a number of others. These, however, do not serve as descriptions on their own; rather, they work by being joined with further phrases or sentences.

Thus we may say that someone desires the death of the King. Here the motive is described by 'desire', and this is supplemented by 'the

death of the King'. This latter phrase specifies the desired (type of) event, or, as we also say, the object of desire. Or again, we may say that someone fears that the King is dead. Here description by the word 'fear' is supplemented by 'the King is dead', which describes the feared situation, that is, the object of fear. And we describe the objects of perception, belief, memory, imagination, hope, expectation, and the rest in a similar way. (So the sense of 'object' here is partly grammatical: the object of desire is what is desired; of belief, what is believed; and so on.)

Now it seems that we understand this use of motive-describing expressions as follows: a descriptive phrase or sentence used in this way serves to assign (or specify) a semantic value for the motive or mental state itself. The value, or semantic correlate, is simply that of the articulating phrase or sentence in question. Thus, schematically, the semantic value of a sentence 'S' is the situation which would render 'S' true; and this, on our mode of description, is the situation which would satisfy the desire that S, verify the belief that S, realize the hope or fear that S, and so on. A phrase or sentence used in this way fits its motive like a glove; the relation of the phrase or sentence to the world seems precisely adapted to specify a corresponding relation on the part of the motive. So such description makes explicit the intentionality, or object-directedness, of motives, representing our minds as engaged with the worldly correlates (or potential correlates) of our words, phrases, and sentences.

Linguistic articulation thus provides us with a powerful tool for psychological understanding. It allows us, in effect, to make a double use of natural language: first in describing the world and then in describing the mind in its engagement with the world as described. This in turn enables us to make use of our grasp of semantic connections for describing and understanding the working of motive, and hence in the explanation of behaviour. This has a number of aspects, two of which can be indicated as follows.

First, as is clear, recycling our worldly phrases and sentences in this way allows us to generate from our relatively short list of words for motives an unlimited variety of psychological descriptions, and in such a way that the burden of information in these descriptions is largely borne by the worldly sentences they embed. The single word 'desire', for example, can be combined with any grammatically appropriate phrase or sentence from natural language (including the vocabulary of physical science), so as to yield a description of any thing, property, or situation which might be an object of desire. So, through such embedding of phrases or sentences, we can specify any describable phenomenon which might move us in this particular way. The same, of course, holds for other ways of being moved; for example, the word 'believe' can be combined with any appropriate sentence to describe any situation which might serve to inform our actions; and so on.

Secondly, this mode of description serves to encode the working of motives linguistically. It enables us, that is, to map causal connections which hold as among motives, or as between motives and reality, by semantic connections which hold for the phrases or sentences in terms of which we describe the motives. This in turn renders the knowledge of semantic connections involved in our understanding of language interconvertible in practice with knowledge of the dynamics of motive. So this use of language gives us, in our commonsense psychology, a sort of natural system for the linguistic – or semantic or hermeneutic – representation of psychological causal role.[16]

To bring this out, let us consider the central and relatively clear cases of desire and belief. As is familiar, we describe desires in terms of their objects or conditions of satisfaction; that is, in an instance of 'Jones desires that S', the conditions in which 'S' would be true are those in which the desire would be satisfied. The semantic correlate of 'S' is the object of desire. But this – the desired action or situation – is precisely what we take it that a desire should bring about (that is, cause) if a person acts on it. If Jones desires that he (Jones) goes to Vienna, then this desire, if acted on, should serve to bring about (cause) Jones's going to Vienna. So the description of a desire in terms of its object is at the same time a description of a cause in terms of an effect which that cause should have, if it operates in a certain way.[17]

In this case the semantic encoding of causal role is so simple and direct that we may not even notice it. It shows, however, in the way our understanding of the sentence which specifies the object of desire already tells us how the desire should operate. In understanding the sentence we use to describe the desire, we know the object or situation which the desire should work to secure. Understanding of language, via semantic articulation, secures (partial) grasp of the role of desire in the production of actions.

Something comparable holds for belief. When we articulate a belief by a sentence, we thereby register that the belief is true in the same situations as the sentence. But also, we take it that the situations in which beliefs are true are precisely those which the beliefs are supposed to reflect or be sensitive to. That is, we take it that beliefs should be formed in such a way as to reflect the facts; and this means

that a belief should be causally sensitive to the circumstances in terms of which it is described. So describing a belief in terms of the conditions in which it would be true is, among other things, a way of describing an effect in terms of circumstances to which it should be causally sensitive.[18] Again, grasp of the semantic values of sentences goes with knowledge of the working of motives, which is so natural and pervasive that we may not even recognize it for what it is.

These roles come together in the simple patterns we use in explaining actions by reference to their reasons. We may, for example, explain someone's climbing a ladder by saying that he wants to reach a shelf and believes that he can if he climbs the ladder. We naturally recognize that such an explanation ascribes to the agent a cogent reason for wanting to climb the ladder; and we can represent this formally – for example, in terms of the symbols we have already used – as follows. The agent

> desires that S_1 (that he reach the shelf)
> believes that if S_2 then S_1 (that if he climbs the ladder, he will reach the shelf)

and so

> desires that S_2 (that he climb the ladder).

The pattern by which we thus represent the derivation of the agent's final desire is comparable to that of a valid argument. In this case the pattern does not relate truth to truth but, rather, truth to the satisfaction of desire. In a valid argument, the truth of the premisses guarantees that of the conclusion. Here, by contrast, the truth of an agent's belief, together with the satisfaction of his derived desire, guarantees the satisfaction of his initial desire. (Hence the validity of the practical pattern emerges if it is read from the bottom up.) So the pattern shows that in forming the desire we want to explain, the agent has sought rationally to ensure the satisfaction of the desire we cite in explanation.

Such patterns track the processes by which belief carries the stamp of reality into the contents of desire, and thence into action. This is a causal matter, but again, one we trace in terms of semantic connection. Here the transfer of information by which the causes of belief shape those of action appears as a pattern of transmission of content, or transfer of semantic value, from motive to motive. The understanding of sentences and their patterns which is part of linguistic competence enables us at once to grasp this transition and to follow in it. So in this case, as in others, in processing our linguistic representations of an agent's thoughts, we in effect think them again (and without effort), and so locate ourselves with respect to him in the cognitive space we share.

Since desires obtain satisfaction through such connection with reality, implicative patterns of this sort have countless significant and connected instances. We read them on to the sequential bodily movements of others readily and pervasively; and such patterns often reach, more or less in series, from overarching goals through the formation of actual plans down to the details of the strings of movements and manipulations of objects in the immediate environment by which an action or project is implemented. We mark the span of such a series of reasons, as it bears on a whole sequence of bodily movements, by such phrases as 'in order to'; thus we might have registered that our agent above was moving a certain way in order to climb the ladder, in order (say) to reach the shelf, in order to get the gun kept there, in order to confront a prowler, in order to protect his person, family, or property, and so on. Thus we hold that an agent desired that S_{n+1} in order that S_n, desired that S_n in order that S_{n-1}, and so on, up through the order.

Each link in such an ordering represents a part, or aspect, of what an agent does – an element in the flow of his or her behaviour – as at once *goal-directed, information-governed,* and *systematically related in these respects to others*. The goals and information thus specified and related pertain to actual or potential features of the environment, taken as semantic values of the phrases and sentences we use in our description of behaviour and motive alike. The order reflects the way in which each goal is *dominated* by those from which it is derived, in the sense that it owes its place in the order at least in part to the fact that it serves as a way or means to them. So the series consists of specifications of causes (motives) which co-operate in action, each making a particular contribution to the order in action in accord with its place in the order of motive. And finally, since we take each goal as something the agent seeks in preference to others, the 'package' of satisfaction described by the logically linked $S_1 \ldots S_n$ upon which we fix in any particular instance represents a sort of maximum – something like the best the agent thought could be managed in the circumstances. Thus each action or aspect of an action which can be interpreted in this way must fit with all others in the strings in which it figures. And the series of interpretations that we fix in the case of each string must also cohere with those that we fix in others, as an intelligible part of the agent's plans and projects and an expression of his or her values over all.[19]

These requirements for interpretive coherence partly parallel those mentioned in connection with the interpretation of codes above; and since the strings in question are interpreted by sentences related by implication, it is plausible to suppose that further constraints are involved. So although these considerations concern only relatively familiar and superficial features of commonsense psychology, they already suggest that we should regard the explanation of action by motive as similar in important respects to the interpretation of language.

We have seen how commonsense psychology articulates motives by sentences whose semantic connections (with one another and with reality) systematically reflect a range of causal connections (with one another and with reality) which hold for the motives themselves. Commonsense psychology thus supplies, via its use of language, what we might call a *causal semantics* for behaviour: that is, an assignment to elements and sequences in behaviour of (motive-articulating) sentences, and thence an assignment of objects (semantic values).[20] We thus assign to each behaviour or sequence which we interpret a set of interrelated (and so mutually constrained) relations to objects, reflecting the information relating to the environment by which that element of behaviour is driven.

In this perspective the interpretation of both language and behaviour ultimately consists in the fixing of semantic values, so chosen as to reflect causally relevant goals and sources of information. The establishing of such causal/semantic links plainly makes for a form of causal explanation; but it should also be regarded as semantic, in the full sense considered in connection with the hermeneutic interpretation of language in I above. That is, in both cases we understand (interpret) by (1) discerning sequences or strings of elements in behaviour, and (2) assigning to the elements and strings relations or sets of relations to objects. The assignments in both cases are systematic, and interrelated by grammatical or logical patterns in accord with which we take the sequences to be generated; in both cases these patterns require that the assignments to each element in a string cohere with those to all others and that assignments made for one string fit with assignments made for others. In interpreting speech, we assign the objects as it were in causally neutral abstraction, in accord with our conception of meaning; then, in interpreting action, we use these interpreted sentences as templates for assigning the objects again, now as objects of desire, belief, or other motives, and so as specifying systematically related causes. But interpreting speech is also interpreting action, so the circle closes on itself: we can take

it to consist throughout in the assignment of values to elements and sequences in behaviour, constrained now by the grammar of language, now by the 'grammar' of action, and so ultimately by both.[21]

It follows, I think, that we should expect the explanation of human behaviour by motive to share the methodological features of the interpretation of language sketched above. So let us return to some of these.

We noted above that many sentences are at once meaningful and novel. So in understanding such sentences, we regularly grasp connections between sentences and their conditions of truth which we do not take as based on simple inductions (inductions not mediated by some complex theory) over past sentences, past pairings of sentence and context, and so forth. And since we also appreciate the new implications of new sentences, the same holds for connections between sentence and sentence. We take it, rather, that our capacity to link sentences with each other and with the world requires to be characterized theoretically, and in terms of the kind of generative structures revealed by, or hypothesized in the course of, interpretation. So if someone were to argue that sentences could not be regarded as having implications or truth-conditions except where these had been established by simple induction, this would be an inductivist fallacy, and one of a particularly destructive kind. For the consequence of adherence to such a fallacy would be the denial of normal understanding of language – the denial of our capacity to interpret one another, as we clearly can.

We can also see that a similar point holds for motives. We constantly form desires and beliefs which are new, in the sense that their articulation requires sentences which are novel; and likewise for the actions to which these motives give rise. So just as it would be a fallacy to hold that semantic connections – linking sentence and sentence or sentence and reality – must be established by simple induction, so it should also be a fallacy to hold that the parallel causal links between motive and motive or motives and their worldly conditions of truth, satisfaction, realization, and so on should have to be established in the same way. And just as adherence to such an inductivist fallacy for language would abrogate linguistic understanding, so adherence to such a fallacy for motive would curtail our natural understanding of action.

Such fallacy would also ignore a flexibility (or creativity) which seems essential to the function of language and action. It is hard to see how our sentences would enable us to communicate and co-operate so fully with respect to our complex and changing world, or how

our motives would keep us going so successfully in it, if they lacked the kind of rule-described capacity to adapt to new circumstances which we are considering. Surely, as our discussions of articulation and interpretation imply, we should regard our capacities for forming sentences and motives as closely connected. New sentences would seem to arise from new desires (to say things, to convey new information), and these from new circumstances; so novelty in speech can be seen as an instance of novelty in action. But any rational action will still flow from a series of logical combinations of sententially articulable motives; so this flexibility remains bounded by the system which seems to make it possible.

III

We have seen something of how the commonsense linguistic encoding of causal role turns the full resources of natural language to the service of specifying similarities and differences among motives (causes of behaviour) and the full synthesizing and projective powers of linguistic understanding to grasping the explanatory import of these specifications. As a result, we are in effect naturally able to describe the mind (or brain) as a semantic engine; that is, a device which draws on information from the environment and transforms this into motive with environment-directed conditions of satisfaction.

The achievement implicit in this mode of description seems striking enough to dwell on for a moment. Any psychology which deals with the way living creatures represent and interact with the world must perforce contain both a vocabulary for describing representations and their transformations and one for describing the way in which such representations and the behaviour to which they give rise relate to the world. These are formidable requirements on description, and it is not easy to think of a more simple, direct, or efficient means of meeting them than that already found in common sense.

Articulation enables us to use worldly linguistic representations directly in the explanation of behaviour and to specify the psychological inputs, transformations, and outputs most relevant to understanding in terms of truth, reason, and the satisfaction of desire.[22] Moreover, the parameters thus described – our worldly goals and sources of information – seem central to any understanding of the mind or brain. So our everyday descriptions and their accompanying norms of function marked in terms of meaning and logic already

provide a remarkably full and detailed framework for further research (although one which is of course subject to modification and enrichment).[23] It seems that here, as in many other cases, nature has provided a solution to problems of information processing which we do well to understand, acknowledge, and use.

Freud was, so far as I know, the first psychologist not only to employ, but also systematically to extend, this flexible, powerful, and natural system of understanding. I have argued elsewhere that his extension is grounded in a mode of inference which has the potential power, scope, and cogency to give strong support to his basic claims.[24] Such strengths in Freud's reasoning have, however, gone largely unacknowledged in the analytical tradition in philosophy. This, I think, is partly due to methodological assumptions whose flaws we are now in a position at least partly to expose. So I want now to consider these, in light of some of the norms of reason which we take good science to embody.

For this purpose, let us continue to use the idea that scientific theories seek to provide the best explanation for the data they cover, and gain support through their ability to do so. Accordingly, we can represent some connected ideals of scientific objectivity, communication, and rational criticism by the following two principles:

1 An explanatory discipline can be regarded as objective only in so far as the explanations, data, and theory used in it are such that the cogency claimed for them can be appreciated by any rational person who fully understands them.
2 The cogency of an explanatory hypothesis or theory can be properly evaluated only by someone who knows how that hypothesis or theory is used; that is, who understands it, the data it is supposed to explain, and how it does so.

The first of these principles relates objectivity to publicity of evaluation or, what is nearly the same, communicability. To be objective, an explanatory discipline must be *rationally transparent*, in the sense that the data and the theory employed in it and the cogency of the relation of explanation between them are appreciable by anyone sufficiently knowledgeable, thoughtful, and intelligent. The second is closely connected with this, and states that evaluation of a theory or explanation must be *well informed*. According to this principle an evaluator must actually understand the theory in question and know the data it covers, so that he or she can rightly consider the purported explanatory relation between them. Both these ideas seem intuitively

plausible; and I think we can see something of their combined working in the success of the physical sciences, which is at once intellectual and social.

As accords with the first principle, the objectivity of physical science is shown by the communication of data and theory, which can be appreciated by anyone competent to understand it. Hence the explanations of physical science can readily be evaluated, more or less at first hand, by a community of people apart from those who devise them. These can reach agreement which is both independent and informed and which, therefore, accords with the second principle. And given this, it is rational, and also in accord with that principle, for people who do not themselves evaluate the theories at first hand to defer to others with greater expertise. (Someone who cannot personally evaluate theory and data about black holes, for example, can best form an opinion by consulting experts; and his deference in this is rational, provided he has good reason to believe that their disciplines are objective, and hence subject in general to evaluation which is well informed.)

Simple and basic as it seems, the requirement of well-informed assessment is opposed by a deep and traditional current in the analytical philosophy of science: that which issues in what we may call 'methodological short-cuts'. These are criteria (characteristically called 'scientific' by their proposers) which are supposed to make it possible to judge a theory or hypothesis in abstraction from particulars of its content, the data it is supposed to explain, or how it explains them. Such short-cuts are meant to enable one to evaluate a theory or explanation definitively, while understanding little of its actual working. So they constitute a sort of philosophical premeditated violation of principles like 2 above.

Thus, according to Popper, for example, corroboration is determined via the application of a single criterion, which is formal or logical: the deductive or predictive power of the theory in relation to the data or observations which it purports to explain. Roughly, a theory counts as scientific only in so far as it entails statements about observations which could falsify it; so in consequence, a theory can genuinely (scientifically) *explain* only phenomena which it could be used to *predict*. Thus theories or explanations which do not yield predictions about the phenomena they explain are no part of science proper, but rather, are metaphysical.

Popper's use of this criterion can be seen in his treatment of Darwin, which is in essentials similar to his celebrated critique of Freud. 'It is important', Popper urges, 'to show that Darwinism [by which he means 'the modern forms' of 'Darwinian theory'] is not a scientific theory, but metaphysical,' and that 'it is metaphysical because it is not testable.' Thus, for example, 'Darwinism does not really *predict* the evolution of variety. It therefore cannot really *explain* it.' Darwin does not offer 'the type of explanation we demand in physics.' For, 'while we can explain a particular eclipse by predicting it, we cannot predict or explain any particular evolutionary change.' Considering the best aspects of Darwinian theory, according to Popper, we can say only that it ' "almost predicts" a great variety of forms of life'; whereas 'in other fields its predictive or explanatory power is still more disappointing. Take "adaptation". At first sight natural selection appears to explain it, and in a way it does, but it is hardly a scientific way.'[25]

Here Popper's criterion enables him to deduce briefly and decisively that Darwin's theories and explanations are only pseudo-scientific, and to do so solely on the basis of predictive power. (And it is clear that the same could have been done, on the same grounds, by someone otherwise entirely ignorant of Darwin's work.) Such a treatment, however, is clearly inadequate. In fact, Darwin's explanations are paradigms of scientific thinking, which should be acknowledged as such on any methodologically unprejudiced account. (I think the same of Freud's, but do not wish to use this in a premiss of the argument.) So Popper's short-cut must be regarded as fallacious.

We should note, however, that Popper was right to assimilate Freud and Darwin and to contrast their work with physics. For both produced accounts of phenomena which had not been explained naturally before and which were similar in structure. (Roughly, Darwin showed that many features of plants and animals could be explained by regarding them as derived, by processes which he was the first to specify, from earlier, sometimes hypothesized forms which had not previously been taken as their origins. The same can be said of Freud, with 'mental life' substituted for 'plants and animals'.) To do this, both had to accomplish what seems the most basic task of explanation: that of exhibiting the data as instances of the generalizations and causal processes which were to provide explanatory information about them. Accordingly, their main work was twofold: on the one hand, they described and specified the new modes of derivation; on the other, they linked the phenomena to these modes by showing repeatedly and in a wide variety of cases that they were as they would be had they been derived as hypothesized.

Popper's short-cut thus applies alike to Darwin and Freud, partly because their research was particularly basic. They worked with new kinds of hypotheses, which did not enjoy the kind of built-in descrip-

tive overlap with the language of observation which we find in physics; so they had to understand and formulate connections between observation and theory case by case. In work of this sort, predictions are often of the delimited kind we saw above in the case of interpretation, in which explanatory success in one instance or area gives grounds for holding that others can be understood in the same way. So both Darwin and Freud claimed support for their hypotheses mainly on the grounds that they served to provide good explanations for the observations they had accumulated and could be relied on to cope with more. And Popper seems to have formulated his criterion precisely to disallow such claims to support in the first place.[26]

So explanatory work like Freud's or Darwin's cannot be dismissed as Popper suggests. Rather, Popper's use of a short-cut criterion to discredit hypotheses in abstraction from their role in explanation is itself to be regarded with suspicion, at least from a rational or scientific point of view. For again, surely, a rational or scientific attitude must involve commitment to fully informed evaluation of hypotheses or explanations. Since this is precisely the process which methodological short-cuts are meant to truncate, there is *prima facie* difficulty in regarding them as respectable. What service could short-cut criteria be relied upon to perform, apart from the perpetuation of judgements made on inadequate grounds?

A further argument may help to bring this out. So far as we have reason to accept a theory if it provides good (best) explanation of the data, then a veridical short-cut would have to exclude the possibility that this is so. To do this, however, the short-cut criterion would somehow have to take account of all the features of theory and data in virtue of which the former could provide good explanation of the latter. (The concepts applied in the criterion should provide something like analyses of our notions of explanation, of what makes one explanation better than another and so on.)[27] But there seems no reason to suppose that any criterion in the short-cut tradition actually does this. So we may expect such criteria to fail in a particular way: they will discount good explanations wrongly, through failing to consider features of explanation which they leave out of account.

This evidently holds for Popper, whose short-cut above requires that all good explanations be assimilable to predictions of particular events. (Cf. his reference to 'a particular eclipse' above and such phrases as 'cannot *predict* . . . so cannot really *explain*', 'predictive or explanatory power', and so on.) This seems a basic mistake: prediction requires certain specific information bearing on the occurrence

of events (time and place); but this is clearly only a small part of the information relevant to explanation and understanding generally. (We understand lightning better for knowing that it is an electrical discharge, but this does not enable us to predict the flashes.) And Popper seems to have left predictive information which is less focused on observable particulars out of account. For, as noted above, an interpretive theory, or our own understanding of language, enables us to predict a number of things, but *not* the occurence or observable specifics of the utterances we use it to understand.

We can see similar failures in the neo-Baconian criteria for assessing psychoanalysis recently suggested by Adolph Grunbaum.[28] Grunbaum urges that 'the establishment of a causal connection in psychoanalysis, no less than in "academic psychology" or medicine, has to rely on modes of inquiry that were refined from time-honored canons of causal inference pioneered by Francis Bacon and John Stuart Mill'. And since he argues that motives are causes, the canons are evidently to apply to claims about them.

Among the 'demands for the validation of causal claims' which these canons lay down are 'the sort of controls that are needed to attest *causal relevance*', which can be satisfied in 'experimental or epidemiological findings'. So, according to Grunbaum, since psychoanalysis is 'replete with a host of etiological and other causal hypotheses, Freud's theory is challenged by neo-Baconian inductivism to furnish a collation of positive instances from *both* experimental and control groups, if there are to be inductively *supportive* instances'.

These quotations suggest that only studies which conform to certain patterns – on which, e.g., a claim to causal connection between a putative cause X and effect Y is established quantitatively, by reference to correlations over X's and Y's in various groups – could provide data supporting psychoanalytic claims about motive. This again would provide a radical short-cut, for it would allow the would-be evaluator to dismiss virtually all observations which analysts have argued are best understood on Freudian hypotheses; and hence, again, to dismiss a theory as lacking support despite ignorance or incomprehension of the explanatory work done by it. (I stress that this is *not* a characterization of Grunbaum's own careful argument, which I have discussed in detail elsewhere,[29] but rather of a use to which the criterion he advocates can well be put.) As with the Popperian criterion, such assessment would not be well-informed in the sense required above, and in particular would fail to take account of the ways in which causal hypotheses can be supported by their role

in explanation. And clearly such assessment would be liable to the methodological errors discussed towards the end of section I, and might provide analogues of the correlational fallacy about motives described in section II.

Short-cut approaches to serious theory can thus be both anti-rational and prone to fallacy. Also, as we can now see, this should be particularly so where they are applied to interpretive thinking, which embodies the kind of information, or is represented by the kind of theory, described above. The cogency of an interpretive hypothesis (and as opposed to alternatives) can become apparent only when it is applied together with a (generally large) family of others, and to a significant critical mass of material. There is, therefore, always the possibility that an uninformed evaluator will fail to use enough of a theory, or fail to use that theory on enough material, to appreciate its explanatory role in full. (Similarly an evaluator can fail to grasp the delimited nature of interpretive prediction, or seek prematurely to impose a superficial correlational grid.) Such mistakes will naturally be more likely, so far as evaluators regard the use of methodological shortcuts as legitimate means of relieving the frustration of incomplete understanding. So we should expect this tradition to be liable in practice to cut against interpretation, and so to obscure what such thinking might otherwise reveal.

Let us now consider the actual place of psychoanalysis in more detail. From what has been said it might seem that the apparent gulf between interpretation and science could be spanned by the devising of interpretive theory. So let us remind ourselves that interpretation is practical, and that practice shows great and apparently ineliminable gaps that remain. In natural practice we are able to *perceive* meaning in the bodily movements (including vocalization and facial expression) of others. Moreover, we can foresee no substitute for such perception; for it results from the processing of data of which we are frequently unaware and which we can envisage no other means of marshalling or treating. So just as sight provides a basic, ineliminable way of learning how things look, acquaintance with others provides a basic, ineliminable way of learning what they think and feel; and we should expect this to remain so, whatever theories we produce.

Psychoanalytic practice – with its emphasis on free association, full disclosure, and uninhibited expression of feeling – is structured to maximize the analyst's acquaintance with material relevant to understanding the analysand in this way, often over the course of years. Although observations made in this way are real and informative, they

are hard to register in full, and impossible to gloss or abbreviate in a way that is fully adequate for communication. (Also, as we have seen, such observations are not connected with the generalizations of the theory which explains them by built-in descriptive links, but rather have to be subsumed anew in each case by interpretive inference. So the terms of the theory do not, in this case, serve so directly to summarize the data they cover.) In consequence, data and theory of this kind cannot be made to travel well; in particular, they are difficult to convey to those who have not made comparable observations.

This restriction is important, since scientific communication ordinarily serves to amplify the voice of reason, by systematically focusing on those inferences in which, as Freud put it, the data speak to one. Hence, in so far as a discipline cannot make such inferences and data readily and fully available, it cannot manifest its objectivity in the way normal for science. A discipline can be fully transparent in principle, but so constrained by the exigencies of actual communication as to seem relatively opaque in practice. In this situation intellectual success within the discipline cannot easily run parallel with fully informed acceptance outside; and those not actively engaged with the full range of data must exercise a kind of trust which is not needed elsewhere. As a result, outsiders, confronted with data which are relatively impoverished and for which they are often well able to frame alternative explanations, will not regard their deference as bound by reason. It thus seems that psychoanalysis can at best be like *science slowed down*, by the need for those considering data and inferences constantly to have recourse to further actual instances of these, as they emerge in practice. Hence, as we observe, criticism of psychoanalysis has often to be set aside as uninformed; but also, advocacy of it cannot carry the authority of fully communicated science.

This flows from the subject itself – from the nature, as it were, of our instruments of mutual understanding – and reflects no fault in the thinking of Freud or his successors. We cannot expect all we learn to be communicable in the same way; and it would be as irrational to confine thought to what can be communicated in a particular way as to use only what came wrapped in plastic. Still, it suggests that the findings of psychoanalysis require to be met not only with the sensitivity to meaning stressed above, but also with a critical alertness, and a willingness to make allowances, appropriate to their interpretive nature. Such complexity of attitude is not required for an approach to normal science.

NOTES

[1] This paper continues a discussion of psychoanalysis and methodology developed from Wollheim's and my seminars and published in 'Epistemology and Depth Psychology: Critical Notes on *The Foundations of Psychoanalysis*' in Clark and Wright (eds), *Mind, Psychoanalysis, and Science* (Blackwell, Oxford, 1988), and 'The Interpretation of Dreams', in J. Neu (ed.), *The Cambridge Companion to Freud* (Cambridge University Press, Cambridge, 1992). There is also related material in 'Synthesis in the Imagination: Psychoanalysis, Infantile Experience, and the Concept of an Object', in J. Russell (ed.), *Philosophical Essays in Developmental Psychology* (Blackwell, Oxford, 1987) and in my Introduction to *Philosophical Essays on Freud*, edited together with Wollheim (Cambridge University Press, Cambridge, 1982).

[2] The debts in what follows to the work of Wittgenstein and Davidson are, I trust, obvious and clearly marked. Also, I was vaguely aware when I wrote this material that I had been influenced by Haugeland, *Artificial Intelligence, The Very Idea* (MIT Press, Cambridge, Mass., 1985) which I had read a few years before. On re-reading ch. 3 of that discussion, however, I see that the debt was more pervasive than I had realized.

[3] For ease of exposition I am using quotation marks in a way Tarski discusses, criticizes, and replaces in the article cited in note 9 below. Also, speaking of knowledge of truth-conditions in this way involves a high degree of idealization, which obscures the role of pragmatics, etc.

[4] Philosophers disagree about what is encompassed in this relation. Davidson and Strawson, for example, take it that sentences correspond to no entities beyond the objects and events referred to by singular terms, whereas others assume that predicates may correspond to further entities (properties), and that sentences can correspond to facts, situations, or states of affairs, which are 'complexes' of objects and properties. Ruth Garret Millikan provides a carefully qualified and thoroughly naturalized version of such an account in *Language, Thought, and Other Biological Categories* (MIT Press, Cambridge, Mass., 1984); and her discussion constitutes a powerful argument that some such account is required. I take this for granted in much of what follows; but my argument would survive reconstrual into any satisfactory account. On these questions, see also note 16.

[5] This distinction is not sharp, for theory can be more or less explicit and also more or less internalized.

[6] See for example Wittgenstein, *Tractatus* (Routledge, London, 1961), 2.1–3.1413, 4.01–4.041; and *idem, Philosophical Grammar* (Blackwell, Oxford, 1974,) pp. 69, 104, 187–90.

[7] This seems to be part of the claim in *Tractatus*, 4.014.

[8] *Tractatus*, 4.002.

[9] For the original rigorous formulation of the idea of a theory of truth, see A. Tarski, 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics* (Oxford University Press, Oxford, 1956). There is a particularly lucid and readable account of Tarski's work in W. V. Quine, *Philosophy of Logic* (Prentice-Hall, Englewood Cliffs, N. J., 1970). It was Davidson's idea to use such a theory of truth as the basis of a theory of meaning, and this requires subjecting a Tarskian theory to constraints which make it reasonable to regard its theorems as yielding correct interpretations. My remarks about Davidson's project are meant to be introductory and not complete. For the full account, see Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford, 1984), and 'The Structure and Content of Truth', *Journal of Philosophy* (January 1990). (For discussion of constraint, see Davidson's 'Reply to Foster' in *Inquiries*.) Also there is a recent discussion of this and a number of related issues in Bjørn Ramberg, *Donald Davidson's Philosophy of Language* (Blackwell, Oxford, 1989).

[10] Tarski's own approach dispenses with the *Tractatus* idea of taking predicates as referring to properties and relations, and also that of regarding sentences as correlated with facts, or 'complexes' of objects and properties. Rather, predicates are related to the objects of which they are true by clauses stating their conditions of satisfaction, and these in turn yield statements of the truth-conditions of sentences. Still, a theorist who thinks it important to acknowledge the role of properties, spatio-temporal 'complexes' of properties and objects, and so on can frame a theory in the Tractarian way. So the question is one upon which further considerations may be brought to bear. (But note that Millikan's *Language, Thought, and Other Biological Categories* provides many qualifications to such Tractarian ideas; see esp. ch. 6.)

[11] I owe this quotation (and my other references to Turing) to Justin Lieber, *Invitation to Cognitive Science* (Blackwell, Oxford, 1991), which appeared just as this essay went to press. Turing's statement is from 'Intelligent Machines', in Meltzer and Mitchie (eds), *Machine Intelligence* (Edinburgh, 1969). Davidson in effect compares his and Turing's views on interpretation in 'Turing's Test', in Said *et al. Modelling the Mind* (Oxford University Press, Oxford 1990).

[12] For a recent discussion of this idea, see P. Lipton, *Inference to the Best Explanation* (Routledge, London, 1991).

[13] Chomsky, 'Review of Skinner's *Verbal Behaviour, Language*, 35 (1926–58).

[14] On the view I am taking here, a true sentence, or again a complex description satisfied by something, will have a semantic correlate of its own, apart from the correlates of the words which compose it; whereas a false sentence or an empty description will not. Still, someone who un-

derstands the words which make up a false sentence or an empty description will know (by knowing the correlates of its components) the kind of correlate it would have if it were true or satisfied. This also counts as knowledge pertaining to semantic values.

[15] The term 'articulation' is Wittgenstein's. See for example his *Philosophical Remarks* (Blackwell, Oxford, 1975), p. 70: 'I call only an *articulated* process a thought . . . Salivation, no matter how precisely measured, is *not* what I call expectation.' Russell called such articulated motives 'propositional attitudes' in his Introduction to the *Tractatus*, and this phrase is still in use. But Wittgenstein's intention there was to relate them, as here, to the worldly states of affairs which would render the articulating sentences true, and not via any abstract intermediary.

[16] Although I am arguing for this view on mainly conceptual grounds, it would suggest that mastery of commonsense psychology should more or less come with mastery of language, the final stage being the use of motive-ascribing words themselves. This empirical consequence seems roughly borne out by recent research, which suggests that children of three years are already adept at belief–desire explanation. See H. M. Wellman, *The Child's Theory of Mind* (MIT Press, Cambridge, Mass., 1990).

[17] A number of explications of the 'should', 'certain way', etc. here are possible (optimal functioning, etc.). I think the best is via Millikan's notion of 'proper function'. The proper function of desires is to produce (cause) their conditions of satisfaction, and the accurate mapping by beliefs of the states of affairs they are about is required for this. See Millikan, *Language, Thought and Other Biological Categories*, esp. ch. 6.

[18] This causal sensitivity is of course a highly complex matter. The more straightforward cases would be provided by beliefs formed on the basis of perception or testimony.

[19] To indicate some aspects of this very roughly: we tend to find agents consistent in so far as strings are characteristically headed by S which are similar or cohering in content (meaning); consistency also requires that S should not change in domination relations from string to string, either *vis-à-vis* other S in a string or *vis-à-vis* alternative S which were candidates for places in a string. (Such changes thus form *prima facie* candidates for change of mind.) Also we take it that the values of the S-specified goals in series are discounted in accord with such likelihoods as the agent attaches to the truth of the conditionals linking them.

These and other constraints mean that in commonsense psychology ascriptions are in effect repeatedly and richly cross-checked, both against one another and against behaviour, in the course of developing an account of a person's motives, and hence form, in practice, a mutually interlocking and supporting system. The same holds for psychoanalytic

ascriptions, for these mainly ascribe values to an agent's motives on the basis of wish-fulfilling representations (phantasies) caused by those motives, and so yield further and deeper specifications of objects of desire, which are checked both against those which arise in understanding action and also those which arise from the interpretation of other wish-fulfilments. This is discussed in more detail in 'The Interpretation of Dreams', cited in note 1 above, and also in the other papers cited there.

[20] The explanations we give of human action are thus, despite their complexity, ultimately comparable to those we might give of the behaviour of honey-bees, by relating their activities systematically to the parameters by which we interpret the 'language' of their dances.

[21] At a more general level, on this view, interpretation consists in linking structures in language, behaviour, and reality which are salient partly because non-coincidentally related. One sees something of this in the way that Augustine has the infant play the discernible orders in behaviour (including speech) and reality off against one another, so as to interpret the former in terms of the latter, in the passage with which *Philosophical Investigations* begins.

[22] I take perception as encompassed in this mode of description, since to perceive that S is generally to have reason to believe that S which is caused by the semantic value of 'S'.

[23] In particular, commonsense psychology can be taken systematically to describe and relate the kind of phase-spaces – the 'as the world presents itself' space and the 'as my body should be' space – which neurocomputational physiologists hope to use to understand the working of the brain. For description of such work, see P. Churchland, *Neurophilosophy* (MIT Press, Cambridge, Mass., 1986), ch. 10; the quoted phrases occur on p. 428.

[24] I discuss Freud's mode of inference with respect to wish-fulfilment in some detail in 'The Interpretation of Dreams', cited in note 1.

[25] These quotations are from pp. 135–8 of section 37, 'Darwinism as Metaphysics', of Popper's contribution to P. A. Schilpp (ed.), *The Philosophy of Sir Karl Popper* (Open Court, La Satle, Ill., 1974).

[26] See for example his dismissive remarks about explanatory power in the title essay of *Conjectures and Refutations* (Routledge, London, 1963).

[27] Popper should be credited with full awareness of this point, for much of his work can be interpreted as trying to prove that it was so.

[28] Grunbaum, *The Foundations of Psychoanalysis* (University of California Press, Berkeley and Los Angeles, 1984); quotations from pp. 47, 128, 185, 189, 280 respectively.

[29] For a fuller discussion of his argument, which I think flawed by reliance on this criterion, see 'Epistemology and Depth Psychology . . . ' cited in note 1. Footnote 21 of 'The Interpretation of Dreams', also cited there, contains material on Grunbaum, but he has informed me that it miscon-

strues his views on the application of Millian reasoning to motives. Therefore much of the argument of that note, and in particular that of the seventh paragraph, does not apply as intended.

[30] I should like to thank Mark Sainsbury and Gabriel Segal for comments which both suggested improvements and saved me from errors.

2

The Nature and Source of Emotion

SEBASTIAN GARDNER

As we said in the beginning that the act of envy had somewhat in it of witchcraft, so there is no other cure of envy but the cure of witchcraft.
Francis Bacon, 'Of envy'

(1) The aim of this paper is to outline and defend a suggestion about the nature and explanation of emotion. The suggestion is, in brief, that emotion is a kind of mental state which cannot be understood apart from – for the reason that it is derived from – the kind of mental state that psychoanalytic theory refers to as phantasy.[1]

The claim that emotion is derived from phantasy is evidently not a commonsensical idea, and it may sound initially more empirical than conceptual. The first part of the argument therefore consists in an attempt to identify, on philosophical grounds, conceptual pressures that might incline us towards accepting it. The second part of the argument seeks to show, with reference to the developmental theory of Melanie Klein, in what ways psychoanalytic theory supports and promises to fill out the philosophical account of emotion argued to be required.

I will begin with some observations about the concept of emotion in commonsense psychology, and define the precise philosophical problem set by emotion which will lead the enquiry.

(2) We have basic working notions of appropriateness for emotions. An instance of anger may be unwarranted if it lacks grounds or if its strength is out of proportion to the grounds that it has. In a similar vein, we can assess emotions in terms of consistency: if an