

Cognitive Neuropsychology and the Philosophy of Mind

TONY STONE AND MARTIN DAVIES

- 1 *Introduction: Philosophy of Science and the Nature of Philosophy*
 - 1.1 *Three Views on Philosophy and Psychology*
 - 2 *Cognitive Neuropsychology*
 - 2.1 *Theoretical and Methodological Assumptions*
 - 2.2 *Examples: Double Dissociation and the Dual Route Model*
 - 2.3 *Metatheoretical Questions*
 - 3 *A Case Study: Recognition without Awareness*
 - 3.1 *A Recognition Impairment*
 - 3.2 *Face Matching*
 - 3.3 *Name Categorization*
 - 3.4 *Learning*
 - 4 *The Myth of Mental Process*
 - 4.1 *The Lost Briefcase*
 - 4.2 *Three Responses*
 - 4.2.1 *First Response: Failure of Imagination*
 - 4.2.2 *Second Response: Conceptual Confusion*
 - 4.3.3 *Third Response: Changing the Subject*
 - 4.3 *Two Explanatory Styles*
 - 4.4 *Independence or Interaction*
 - 5 *Philosophical Lessons from Cognitive Neuropsychology*
 - 5.1 *The Modularity of Mind*
 - 5.2 *The Refinement of Mind*
 - 5.3 *The Elimination of Mind*
 - 5.4 *Consciousness*
 - 6 *Conclusion*
-

I INTRODUCTION: PHILOSOPHY OF SCIENCE AND THE NATURE OF PHILOSOPHY

It is sometimes argued that much of the best new work in the philosophy of science is occurring where philosophers engage themselves with specific research programmes in, for example, quantum mechanics, evolutionary

biology, or cognitive science. Our remarks in this paper will exemplify the trend recommended by those arguments. We shall focus our attention upon one aspect of current research in cognitive psychology, namely cognitive neuropsychology.

We shall not, however, be primarily concerned with what we can learn about the methodology, and core theoretical concepts, of this particular research programme. Of more importance is what the pattern of philosophical engagement may reveal about the nature of philosophy itself.

1.1 *Three Views on Philosophy and Psychology*

(1) Someone might say that philosophy, having no special subject matter of its own, is a metadiscipline that makes its living by picking up the conceptual crumbs from the tables at which other disciplines feast. Relatively new disciplines, like cognitive psychology, can certainly be relied upon to provide a rich diet of conceptual morsels. Psychology's explanatory repertoire includes such tasty items as information processing, mental representations, and tacit knowledge. So there is no shortage of philosophical work to be done at this metatheoretical level.

On this first view of the nature of philosophy, 'philosophy of science is', as Quine puts it, 'philosophy enough' (quoted by Cummins [1989], p.v). Those who adopt this view regard the areas of philosophy whose practice appears to be unmolested by science as living on borrowed time. Just as the development of physics, chemistry, and biology in the seventeenth, eighteenth, and nineteenth centuries gradually pushed back the range of philosophical activities, so the development of the science of psychology in the twentieth century will eliminate the philosophy of mind, replacing it with the metatheoretical philosophy of psychology.

(2) There is, of course, a different view of our discipline, according to which its core is quite autonomous from the sciences (and from such disciplines as history, sociology, and literature). On this alternative conception, philosophy is a pure *a priori* discipline which, by the careful examination of our concepts, investigates problems that have a perennial fascination. The philosophy of science is then the relatively peripheral part of the discipline in which these same *a priori* skills are let loose upon the proprietary concepts of particular sciences, instead of the everyday concepts of our common-sense scheme.

The evaluation of the *a priori* claims of philosophy requires a fertile source of *a priori* possible counterexamples. And the discoveries of science may sometimes make good a failure of philosophical imagination: the actuality of a counterexample rendering vivid its *a priori* possibility. But on this conception there can be no essential interdependence between science and the core of philosophy. If a philosopher makes a claim that appears to be amenable to empirical investigation, then that simply shows that he has overreached

himself. Philosophers should be ever vigilant to ensure that extraneous empirical material is kept out of their work.

Indeed, it might be argued that one way in which philosophy progresses is for this empirical detritus to be handed over to the relevant empirical discipline. Isaiah Berlin presents a picture of the development of philosophy not unlike this when he says ([1980], p. 5):

The history of thought is thus a long series of parricides, in which new disciplines seek to achieve their freedom by killing off the parent subject and eradicating from within themselves whatever traces still linger within them of 'philosophical' problems, i.e. the kind of questions that do not carry within their own structure clear indications of the techniques of their own solution.

The history of philosophy, on this account, is of a gradual winnowing down of the discipline to the essential philosophical (that is, non-empirical) questions.

(3) A third view can be mentioned quite briefly. This casts philosophy in a critical role, revealing conceptual confusions at the very heart of ostensible disciplines such as cognitive psychology. (See *e.g.* Baker and Hacker [1984], for an expression of this view, particularly in respect of that part of psychology that constitutes theoretical linguistics.)

While we do not propose this primarily critical role for philosophy *vis-à-vis* psychology, we also share neither the view that philosophy is properly restricted to metatheoretical philosophy of science, nor the view that the true domain of philosophical enquiry is hermetically sealed off from empirical infection. On the contrary, we regard appearances of interaction between philosophical and psychological theories as *prima facie* veridical. By taking these appearances at face value, we hope to learn something about philosophy, and about the philosophy of mind, in particular.

Our plan in this paper is as follows. First (Sections 2 and 3), we present a more detailed account of our chosen empirical research programme: cognitive neuropsychology. We set out some of its theoretical commitments (Section 2), and describe a case study (Section 3). Second (Section 4), we argue for the interactive conception of the relationship between philosophy and psychology, by considering the neo-Wittgensteinian claim that there are no such things as mental processes. Third (Section 5), we suggest some general directions that further interdisciplinary research might take.

2 COGNITIVE NEUROPSYCHOLOGY

Cognitive neuropsychology is the branch of cognitive psychology in which models of, or theories about, normal cognitive processes are developed, evaluated, and refined, in the light of data provided by the empirical investigation of people with acquired disorders of cognition (see Ellis and

Young [1988]; Shallice [1988]). Although this branch of the subject may not be very familiar to philosophers, it is not some mere esoteric backwater. On the contrary, 'virtually all major aspects of cognition are being studied from the perspective of cognitive neuropsychology' (Coltheart [1985], p. 3).

The cognitive neuropsychologist aims to investigate our normal mental states and processes—for example, memory, reading, visual object recognition—by examining the breakdown of those processes due to such traumas as strokes and closed head injuries.

2.1 *Theoretical and Methodological Assumptions*

The recent flowering of this research programme has been based upon a number of theoretical and methodological assumptions. Here, we briefly mention four.

(1) First, cognitive neuropsychology insists upon the importance of interaction between the investigation of normal and damaged cognitive processes. The aim of accounting for deficits apparent in patients with acquired cognitive disorders is most fruitfully pursued if it is based upon the development of models of normal processing. In turn, data from patients are important in the evaluation and development of these models. In particular, if a pattern of selective impairment and preservation is observed which, according to some model of normal processing, could not arise, then that counts as evidence against the model. Ideally, what is sought is converging evidence from studies of normal and brain-damaged subjects.

Models of normally functioning cognitive processes should provide theoretically motivated accounts of acquired disorders. Typically, the account is in terms of damage to, or abolition of, certain component processes, while others continue to operate normally. The prospect of this type of explanation rests upon a second theoretical assumption.

(2) This second assumption is that the mind is modular. There are discrete, task-specific, processing components, that are wired up in such a way that they may be selectively impaired. Furthermore, brain damage does not lead to a massive reorganization of the prior modular structure: damage to components leaves the operations of the unimpaired modules unchanged. This latter part of the modularity assumption is what Caramazza ([1986], p. 52; cf. [1984]) calls the *transparency assumption*.

(3) The third assumption—this time methodological—is that cognitive neuropsychology should proceed by means of single case studies. There is not much to be learned from group studies in which data are averaged over patients who present similar gross symptoms. Correlatively, it is claimed that the nineteenth-century neuropsychologists' notion of a syndrome—a co-occurring constellation of symptoms—is no longer useful. (On this issue, see

the main argument of Caramazza [1986]; and for a dissenting view, see Bub and Bub [1988]. For discussion, see Shallice [1988], pp. 203–12.)

(4) Fourth, cognitive neuropsychology reckons the neurophysiological details of patients to be more or less irrelevant. The assumption is that the study of the diseased or injured brain, as such, can provide little or no help in the construction of psychological models of either the normal or the disordered cognitive system. In particular, contemporary cognitive neuropsychology differs from its nineteenth-century ancestor in not being particularly concerned to map specific cognitive functions to local brain areas. Cognitive neuropsychology thus stresses the 'cognitive', rather than the 'neuro'.

This fourth assumption may, in fact, be maintained in more or less dramatic forms. In its most extreme form—as the claim that neurophysiological evidence is in principle irrelevant—it is part of what Shallice ([1988], p. 203) labels *ultra-cognitive* neuropsychology. However, virtually all cognitive neuropsychologists agree that—in accordance with Marr's [1982] hierarchy of levels—psychological theories are constrained from below by the facts of neurophysiology. (Theories at Marr's level two—the level of the algorithm—are constrained from below by theories at level three—the level of hardware implementation, and from above by theories at level one—the level of the computational task that is to be performed.) Consequently, the claim that neurophysiological details are not especially relevant is usually pragmatically grounded. As Shallice himself says ([1988], p. 214): 'To hope for an advance in theories of the functional organisation of cognition by paying special attention to issues of localisation is not, at present, a promising strategy.' The advocate of this moderate form of the fourth assumption need not deny the interest of new methods of neurophysiological investigation, such as positron emission tomography (PET) (*e.g.* Petersen *et al.* [1988, 1990]). But he is likely to maintain that the PET scan studies are guided by an antecedent psychological theory that has already been informed by the findings of cognitive neuropsychological case studies.

These theoretical and methodological assumptions will be seen at work as we proceed to some examples of cognitive neuropsychological arguments.

2.2 Examples: Double Dissociation and the Dual Route Model

Cognitive neuropsychological case studies play a role in the *evaluation* of theories about normal processing. Data from brain-damaged patients may enable us to choose between two competing theories about normal processing. And cognitive neuropsychology can *constrain* the development of theories about normal processing. Sometimes, data from brain-damaged patients simply rule out a whole family of possible theories. Ellis [1987] cites a clear example.

A number of patients have been studied who are good at reading familiar

words, but poor at reading unfamiliar words and invented non-words such as 'vib' and 'slint'. In the starkest case (patient W.B.) reading aloud of familiar words, including words with irregular spelling, is very good (85 per cent correct), while reading aloud of non-words is very poor (0 correct out of 20). Of this case, Ellis remarks ([1987], p. 403):

This patient is sufficient to disprove any theory that asserts that the identification of familiar written words necessarily involves an early stage in which the spoken form of the word is assembled piecemeal from its written form by the application of grapheme-phoneme conversion rules.

(For the case of patient W.B. see Funnell [1983].)

This is a relatively uncontroversial form of constraint. But there is a second, and more dramatic, way in which data from patients are claimed to constrain the development of theories. In what might be thought of as classic examples of the *logic of cognitive neuropsychology*, strong conclusions about independent modules are drawn from premises about dissociations between cognitive deficits.

Here is a textbook presentation of an argument in this style (Harris and Coltheart [1986], p. 232):

If a theory of normal language processing were proposed in which there were a single processing system responsible for dealing with spoken language—a system-used both for perceiving and producing speech—then one ought never to see patients with intact speech perception and impaired speech production, nor patients with the opposite disorder. The fact that both these forms of aphasia are frequently observed suggests that there are separate systems for perceiving and producing speech.

Here the argument is for the recognition of separate systems, or modules, within the language processing system. But equally, the same style of argument could be used with respect to larger-scale cognitive functions. Thus, from the facts that one can find patients whose language is impaired but who have intact visual object recognition, and one can find patients whose visual object recognition is impaired while they have normal language, the conclusion is drawn that there is a language module and a separate visual object recognition module.

The general form of argument from dissociations is clear enough. The systems X1 and X2 that are responsible for the performance of two tasks T1 and T2 are argued to be independent systems or separate modules, on the grounds that performance of each of the tasks can be impaired while performance of the other remains intact. The argument is from double (that is bidirectional) dissociation of deficits to modularity.

Within at least some parts of the cognitive neuropsychology community, it is held to be important to have evidence of a double dissociation. For suppose

that we found merely that T2 could be impaired with T1 intact. Then, it might reasonably be maintained that the two tasks are performed by a common system, that T2 is a more difficult task than T1, and that brain damage impairs difficult tasks more than easier ones. A double dissociation clearly blocks this line of reply. (See Coltheart [1985], p. 10; but for the view that double dissociations are not evidentially privileged, see Caramazza [1986], pp. 64–5.)

This style of argument can be illustrated further—this time at a finer grain—if we consider the task of reading single words aloud. *A priori*, it would appear that there are two kinds of information relevant to the performance of the task in the case of a word, such as 'mint', whose spelling is regular. One kind of information concerns the letters or letter clusters (graphemes) within a word, and their order. These letters or letter clusters correspond to phonemes, and—because the spelling is regular—those phonemes, together with their order, determine the correct pronunciation of the word. The other kind of information concerns the identity of the whole word. For just as the meaning of a word is stored in memory, so also the correct pronunciation is stored as a word-specific phonological form.

Performance of the task utilizing information about letter-sound correspondences is said to take the non-lexical route. Performance of the task utilizing information about a word-specific phonological form is said to take the lexical route. Evidence for the use of the first kind of information in normal subjects is provided by the fact that they can pronounce non-words like 'slint' and 'vib'. Evidence for the use of the second kind of information is provided by normal subjects' ability to achieve the correct pronunciation of irregular words like 'pint' and 'yacht'.

Thus, one possible model of the reading aloud process in normal subjects would have two distinct and autonomous routes: one route drawing upon information about letter-sound correspondences, the other drawing upon information about word-specific phonological forms. In the case of a regular word, either route would serve. In the case of an irregular word, the second route would be needed. In the case of a non-word, the first route would be required. (For general discussion of the dual route model, see Humphreys and Evett [1985].)

A cognitive neuropsychological argument for two independent routes can be constructed using data from the patient W.B. already mentioned, and another patient M.P. (For patient M.P. see Bub, Cancelliere and Kertesz [1985].)

Recall that patient W.B. presents accurate reading of words, both regular and irregular, but is quite unable to read non-words. Patient M.P. presents the ability to read aloud regular words and also non-words like 'vib' and 'slint'. But M.P. has considerable difficulty reading aloud irregular words, often introducing regularization errors; for example, reading 'pint' to rhyme with 'hint', or

'yacht' to rhyme with 'matched'. In addition, M.P.'s pronunciation of non-words that are made up from irregular words reveals regularization in 39 out of 43 cases (*e.g.* 'mave' to rhyme with 'gave' rather than with 'have').

Thus, there are patients—such as W.B.—who can read aloud both regular and irregular words, but who cannot read non-words at all. And there are patients—such as M.P.—who can read aloud regular words and non-words, but who make regularization errors when reading words like 'pint' and 'yacht'. From this double dissociation, the logic of cognitive neuropsychology leads to the conclusion that there are two separate systems for reading aloud: two independent routes from orthography to phonology.

The two patients just described [W.B. and M.P.] demonstrate a double dissociation . . . that is extremely clear. It would seem, then, that the two patients in particular, and the forms of acquired dyslexia known as phonological and surface dyslexia in general, provide strong evidence in support of the view that the information-processing system we use for reading aloud . . . ought to be thought of as comprising two separate components. (Coltheart [1985], p. 13)

It is worth being quite clear about the nature of the double dissociation in this example. Patient W.B. can correctly pronounce words, but not non-words. But the reverse dissociation exhibited by patient M.P. is not that of being able to pronounce non-words but not words. (Indeed—as Caramazza [1986], pp. 64–5, in effect points out—it is quite unclear what conclusion would be licensed by the discovery of that pattern of performance.) Rather, whereas W.B.'s reading aloud extends beyond regular words to irregular words but not to non-words, M.P.'s correct pronunciation extends beyond regular words to non-words but not to irregular words.

This same pattern of argument from dissociation could be used to support fractionation at successively finer grains. For example, it is possible—'[r]emorselessly pursuing the logic of cognitive neuropsychology' (Coltheart [1985], p. 17)—to argue for further modular structure within each of the two routes for reading aloud. In particular, it is possible to argue for three components to the non-lexical route, and then plausibly to identify failure in one of the three as the source of W.B.'s inability to use that route.

2.3 *Metatheoretical Questions*

These examples, and the style of argument that they illustrate, certainly provide plenty of raw material for the philosopher as metatheorist. It is inevitable, for example, that a philosopher of science will ask whether the notion of modularity that is deployed by the cognitive neuropsychologist is the same as Fodor's [1983] notion.

A philosopher might also ask whether there is not some further theoretical assumption at work in the arguments from double dissociation. For, even given a general assumption of modular structure, all that strictly follows from

a double dissociation between tasks T1 and T2 is that there is at least one component of the system X1 (responsible for T1) that is not also a component of X2 (responsible for T2), and that there is at least one component of X2 that is not also a component of X1. That is a very long way from the claim that X1 and X2 are completely independent systems or modules.

We shall not pursue either of these questions now, though they deserve careful attention. They serve well enough to illustrate the rich diet that cognitive neuropsychology offers to the metatheoretically inclined philosopher of science. But, for the remainder of this paper, our concern lies elsewhere.

3 A CASE STUDY: RECOGNITION WITHOUT AWARENESS

With the basic methodological structure of cognitive neuropsychology clear, let us look at another example in more detail. The example chosen is a case study of prosopagnosia (inability to recognize faces) reported by Edward De Haan, Andrew Young, and Freda Newcombe [1987]. (See also Young and De Haan [1988], Young, Hellawell and De Haan [1988]; Young and De Haan [1990].)

3.1 A Recognition Impairment

De Haan *et al.* [1987] report an empirical investigation of a patient P.H. who, following a closed head injury, was entirely unable to recognize faces, even those of his close family.

He [P.H.] was completely unable to overtly recognise familiar people from photographs of their faces. Of the hundreds of famous faces shown to him during the course of this investigation, he has only achieved spontaneous recognition on one occasion. (De Haan *et al.* [1987], p. 389)

(In fact, subsequent research has found that P.H. can recognize a few faces, although not usually consistently (Young [In press]):

When we first began working with P.H., he seemed to us to be completely unable to achieve overt recognition of familiar faces. De Haan (1987) noted that of the hundreds of famous faces that they had shown him, only one (Mrs. Thatcher) had been spontaneously recognised, and that on only one occasion.

Since then, we have also noted other occasions on which P.H. has overtly recognised a face during the last few years. There are about a dozen faces he has recognised occasionally, but the only face we have noticed is beginning to be fairly consistently recognised is Mrs. Thatcher's.

It remains the case that P.H. has an extremely severe impairment of overt recognition of familiar faces.)

It is important to note that P.H.'s deficit was highly specific. P.H. was able to recognize people from the visual and oral presentation of their names. For example, when asked to sort photographs of faces into the categories FAMILIAR and UNFAMILIAR his success rate was no better than chance,

whereas on the same sorting task applied to written names he was highly accurate (29–32 correct). In fact, quite generally, prosopagnosia can occur without failure of word recognition; and the reverse dissociation is found, too (see *e.g.* Shallice and Saffran [1986], and for some discussion Young and De Haan [1990]).

The relationship between face recognition and visual object recognition is less clear. Failures of visual object recognition seem almost always to be accompanied by problems with face recognition. And, in the reverse direction, prosopagnosic patients usually reveal some visual problems with objects other than faces; for example, they may have trouble distinguishing visually amongst objects of some kind—amongst different makes of car, perhaps. As it happens, there are patients who are claimed to show a clean dissociation in each direction. But, even without a full set of reverse dissociations, it is compelling to suppose that there is some cognitive machinery that is more or less dedicated to the task of face recognition. (See Humphreys and Bruce [1989], p. 88–95 for further discussion.) How might this machinery work?

It is possible to begin with a fairly simple model of the cognitive processing involved in face recognition: we can think of the recognition as proceeding through three different processing levels. At a first level, the face is identified as familiar or unfamiliar. Let us suppose that a structural description is produced which is then matched against a library of familiar structural descriptions. At a second level, what psychologists call *semantic* information becomes available. This is information about properties of the person, such as his or her occupation. Then, at a third and final level, the person is identified and his or her name is accessed.

In normal everyday life, it often seems to happen that we are able to process a face up to only the first or second of these levels. Thus we may recognize a face as familiar, but be unable to add any further information: 'I've seen him before, but I just can't place him'. This would seem to be an example of processing up to the first level only. Then there are cases where one knows that the face is of a German sprinter, but one just cannot remember her name. This would seem to be processing up to the second level. It is only when the processing proceeds up to and through the third level, and access to the name is achieved, that satisfactory recognition is reckoned to have taken place.

The aim of the case study carried out by De Haan and his associates was to investigate whether this three-stage model of the normal processing of faces enables us to give a theoretically illuminating account of P.H.'s disorder. In particular, the question is this. If normal face recognition does take place through a series of successive levels, then which of the levels is damaged in P.H.?

In this study, De Haan *et al.* are not trying to show that the three levels postulated in the model are separable processing modules, as was the case with the case studies mentioned earlier with regard to the dual route model of

reading. In later work (Young and De Haan [1991]), they do argue for the modularity claim; and it is worth noticing that the argument cannot rely on double dissociations, since the model of face recognition is hierarchically organized along a single route. (We cannot expect to find evidence of an intact later stage when an earlier stage is impaired.) But the aim of the study that concerns us here is just to discover whether any of the postulated processing stages are still operative, despite P.H.'s inability to recognize faces overtly.

The experimental challenge is to design experiments that will reveal the intact activity of some of the processing stages, should there be any, without relying upon the patient making any overt recognition judgements. For recall that, as far as explicit reports go, P.H. is not merely unable to put a name to a face. He cannot even reliably say whether a face is familiar to him or not.

3.2 Face Matching

The experimental design is simple but cunning. It is known that the performance of normal subjects on various tasks involving faces can be affected by prior or simultaneous exposure to other information.

It is known, for example, that when normal subjects are required to judge whether two photographs of faces are photographs of the same face or of different faces their speed of decision increases when the photographs are of familiar faces. Assuming that performance of this task requires that the face recognition system must be working up to the level of accessing the stored structural descriptions of the faces of familiar people, the hypothesis is that if P.H.'s face recognition system is intact up to that level, then his pattern of performance on the same tasks will be the same as that of the normal subjects.

The data in Table 1 reported by De Haan *et al.* ([1987], p. 392), for P.H.'s performance on the face matching task, appear to confirm that his face recognition system is intact up to that first level. P.H. performs this task worse

TABLE 1. Mean reaction times and error rates on the familiar and unfamiliar face matching task for P.H. and 16 normal subjects of comparable age. (Data for normal subjects are from the study reported by Young *et al.* [1986]. Identity Matching Condition.)

	Familiar	Unfamiliar
P.H.		
Error rate (%)	18.7	16.4
Mean RT (msec)	2550	2762
Normal subjects		
Error rate (%)	1.6	3.9
Mean RT (msec)	977	1045

than normal subjects: he makes more errors and his responses are much slower. But the advantage of familiar faces over unfamiliar is statistically significant in P.H.'s performance, just as it is for normal subjects—and this despite the fact that P.H. is quite unable to report whether a face is familiar or not.

The claim is then that these data confirm:

- (i) that there is a level of processing at which the stored structural descriptions of familiar faces are accessed (though the experiment does not provide evidence that it is an isolable sub-system); and
- (ii) that despite P.H.'s inability overtly to recognize faces, some degree of covert recognition of the faces does take place.

For how could the familiarity of a face have facilitated performance unless some covert recognition of the face had taken place?

3.3 Name Categorization

Consider now a second experiment. A well-confirmed result in normal subjects is that knowledge about a familiar face interferes with performance on name categorization tasks. So, for example, if normal subjects were asked to categorize the name 'Bobby Charlton' as the name of a sportsman rather than of a politician, they would present faster reaction times when the name is presented on its own than when it is presented alongside the face of someone from a different category—*e.g.* if the face of Neil Kinnock were presented simultaneously with the name 'Bobby Charlton'.

In the name categorization experiment carried out by De Haan and his colleagues [1987]—using the two categories politician and TV personality—there are five different conditions (p. 400).

Same person: The face and name are of the same person.

Name only: Only a name is presented.

Unrelated: The face and the name belong to people from different categories, who do not look particularly like each other.

Semantic category relation: The face and the name belong to different people from the same occupational category, who do not look particularly like each other.

Visually related: The face and the name belong to people from different categories, but who are similar in physical appearance. For example, the face of Frank Bough is paired with the name of Neil Kinnock, and vice versa (see Figure 1).

In this experiment, P.H. presents the same crucial pattern of results as do normal subjects. Table 2 represents the complete table of results from De Haan *et al.* [1987], p. 402). As in the first experiment, P.H.'s responses are much slower overall. But, for P.H., just as for normal subjects, his responses in the

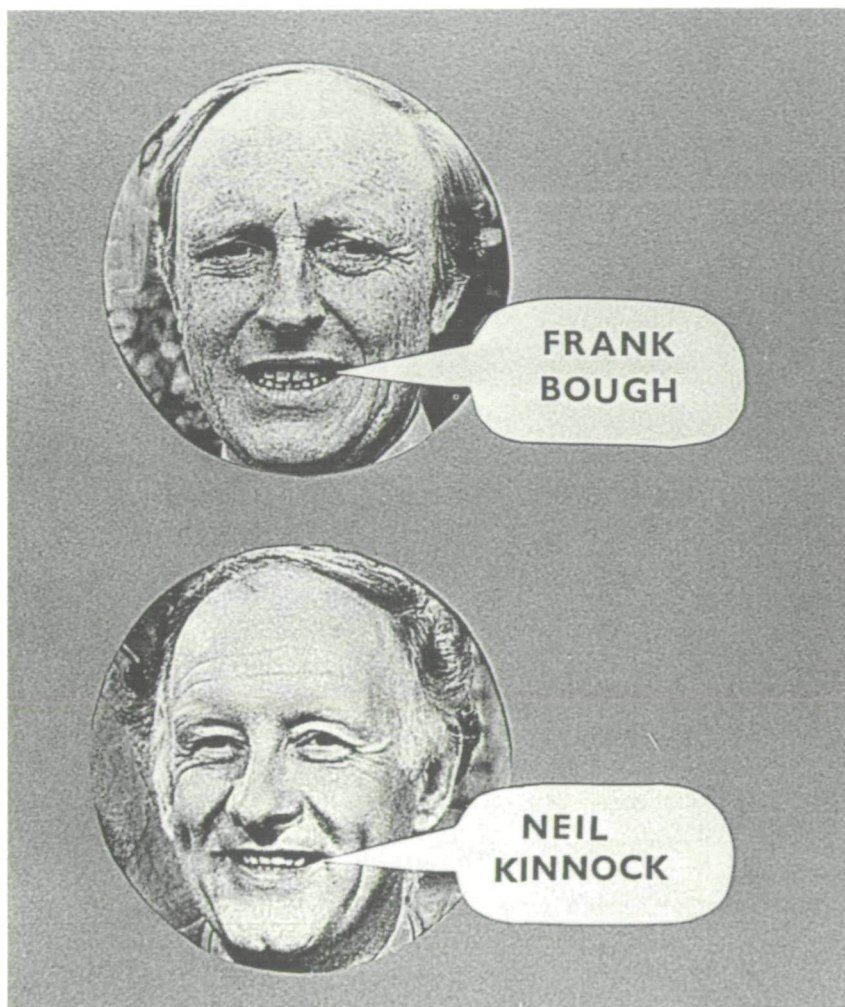


FIGURE 1 Examples of stimuli from the Name Categorization experiment, *Visually related* condition. The face of the TV personality Frank Bough is paired with the name of the politician Neil Kinnock, and vice versa. As can be seen, the two faces are of similar appearance. (Reproduced with permission from De Haan *et al* [1987], p. 401.)

Unrelated and *Visually related* conditions are significantly slower than in the *Name only* condition. (See the difference between the third and second columns, and between the fifth and second columns, in the table.) The face of someone from a different occupational category inhibits the categorization of a name; and this effect persists even when the person whose face is presented has a similar appearance to the person named.

TABLE 2. Mean reaction times and error rates on the politician versus television personality face-name interference task for P.H. and 12 normal subjects of comparable age. (Data for normal subjects are from Young, Ellis *et al.* [1986], Experiment 4: Name Categorisation Condition.)

	Same person	Name only	Unrelated	Semantic category relation	Visually related
P.H.					
Error rate (%)	1.4	5.6	1.4	1.4	2.8
Mean RT (msec)	1506	1433	1604	1484	1598
Normal subjects					
Error rate (%)	0.3	1.4	1.4	0.2	2.3
Mean RT (msec)	746	772	804	774	799

Furthermore, there is no possibility that the interference effect in P.H. is the product of overt recognition of the people whose faces are presented in the experiment. For, when P.H. was asked to categorize the faces (as faces of politicians or TV personalities) he performed at close to chance levels (55.5 per cent correct; chance = 50 per cent correct).

Two claims are made on the basis of this second experiment; namely:

- (i) that there is a level of processing at which semantic information is accessed (although let us stress again that this particular experiment provides no evidence that it is an isolable module); and
- (ii) that despite P.H.'s inability overtly to recognize faces, covert recognition of faces as the faces of people with particular occupations does take place.

For how could the face of Neil Kinnock have inhibited the categorization of the name 'Terry Wogan' as that of a TV personality, unless covert recognition of that face as the face of a politician had taken place?

3.4 Learning

The third experiment we will mention concerns the effect of familiarity on the learning of true and false names of faces that P.H. cannot overtly identify.

P.H. was taught to associate a name ('Terry Wogan', 'Neil Kinnock') with a face. The faces were familiar to P.H. although he was, of course, unaware of that fact. Indeed, in order to rule out the possibility that P.H. had a degree of overt recognition of these people, the faces used in the test were those for which he had chosen an incorrect name in an earlier forced choice face-name matching task.

In half of the cases, the name taught was the pictured person's correct name; in the other half, the name taught was incorrect. The question to be

TABLE 3. Number of correct choices (maximum = 3) of face to match stated names on each trial. The stimulus faces were chosen to conform to the criterion of absence of correct overt identification in a binary-choice screening task.

Trial number	1	2	3	4	5	6	7	8	9	10	11	12
True pairings	2	2	1	1	3	3	2	3	2	2	2	3
False pairings	1	1	1	1	3	2	3	2	0	1	1	2

investigated was whether P.H. would show better learning for true face-name pairings than for false ones. An advantage for the true pairings over the false would, once again, indicate that some early stages of the recognition process are intact in P.H..

The experimenters tested P.H.'s learning by asking him to match faces to six names; for half the names he had been taught a true pairing, and for half a false pairing. As the above data from twelve trials shows (Table 3), P.H. did indeed present evidence of significantly better learning for the true pairings (De Haan *et al.* [1987], p. 406). This finding does not, by itself, show that P.H. has covert access to the name of the person whose face is pictured. It does not, that is to say, reveal that in P.H. all three stages in the recognition process are intact. In an attempt to discover just which stages of processing are intact, some further experiments were carried out using the same kind of learning task.

Young and De Haan [1988] explored access to names and to so-called semantic information about a face. As well as being better at learning true pairings than false pairings in the case of faces and names, P.H. was also better at learning true pairings of faces and occupations—politician, actor, and so on—than false pairings. However, while this effect of better learning of true pairings over false pairings was found for full names and for occupations, it was not found for first names only ('Neil', 'Bobby'), nor for detailed semantic information (political party, or particular sport). The suggestion is, then, that the learning effect is produced, not by covert access to the names associated with faces, but by covert access to the occupations associated with faces together with overt association of occupations with names.

A final point that should be mentioned about this case study is that P.H. shows no awareness of his success. That is, P.H. shows no awareness that he is performing better on some trials than on others. Indeed when informed that he is successful on some of the tasks, he expresses disbelief. This is a phenomenon that occurs in many other cases where perceptual processing without awareness is found. For example, patients who are suffering from the condition

known as blindsight and are able to evidence covert-perception of light flashes in the area blanked out by their scotoma maintain that they are guessing even when told that they are successfully indicating the position of the flash.

That concludes our exposition of the case study. We now turn to philosophical consequences.

4 THE MYTH OF MENTAL PROCESS

Philosophers of a Wittgensteinian persuasion often challenge the very idea of mental processes. We claim (4.2) that this challenge cannot be sustained, save perhaps as a terminological recommendation concerning the term 'mental'. Furthermore, we claim (4.3) that a terminological distinction between the domains of folk psychology (the 'genuinely mental') and subpersonal information processing should not be allowed to suggest the constitutive irrelevance of one domain to the other. Ultimately, of course, we aim (4.4) to motivate our interactive conception of the relationship between philosophy and psychology. In pursuit of these aims and claims, we consider Norman Malcolm, David Hamlyn, and John McDowell.

4.1 *The Lost Briefcase*

First, let us turn to Malcolm's paper 'The Myth of Cognitive Processes and Structures' [1971]. Malcolm argues that claims made by psychologists, that the act of recognizing something—say, the face of some person as the face of a friend—or of remembering something—such as, where I have left my briefcase—involves mental processes, can be seen to be false by the 'consideration of a few examples'.

Thus, he says [1971, p. 159]:

Sometimes we go through a process of *trying* to remember. Suppose that you cannot locate your briefcase. You remember that you were carrying it when you left your office. You review in your mind, or aloud, your itinerary on the way home. 'I walked to the bank and cashed a check. Did I have the briefcase when I left the bank? I'm not sure. I then went to the bookstore and bought an atlas. Now I know that I *did* take the briefcase into the bookstore, for I remember putting it down when I paid the cashier . . .' While saying or thinking these things you may have had feelings of anxiety: images of the streets, the bank and the store may have passed swiftly through your mind; finally when the solution came, you may have had a feeling of relief as if a weight had been lifted from you.

Malcolm's point is that, on any particular occasion when I remember the location of my briefcase, the described mental happenings may go on or they may not. But their absence is consistent with our having a perfectly genuine case of *trying to remember*, or indeed of *successfully remembering*, where my briefcase is. Hence, for Malcolm, the occurrence of such mental happenings cannot be constitutive of my remembering.

These claims are apparently intended to threaten the practice of cognitive psychological explanation; but it is not altogether easy to construe them in a painful way. After all, most cognitive scientists would agree with the specific point that Malcolm makes. There is no evidence that there are *conscious mental happenings that invariably accompany an instance of remembering*. But this is beside the point, for the cognitive scientist who models mental processes is usually not making a claim about conscious mental happenings. On the contrary, the processes in which the cognitive scientist is interested are often quite inaccessible to consciousness: we cannot, for example, introspect the nature of early visual processing or of parsing.

David Hamlyn—who explicitly records his agreement with Malcolm—says, in a similar vein ([1983], p. 304):

I have no model of the mind or model of mental processes to put in place of those which I reject, because I do not think that there is any such model. If we think that there is one, then I think that we fail to learn the lesson which Wittgenstein tried to teach when he said what he said about the myth of mental process.

What we seem to have here is an extension from Malcolm's specific example to a quite general claim (a claim with which Malcolm, himself, would agree). On the face of it, the claim is that when someone remembers something or recognizes someone there is nothing that invariably goes on and that deserves to be called a mental or cognitive process.

On one construal, this position may be seen by a cognitive scientist as involving an overambitious claim that turns out to be empirically false. Suppose that the putative explanation of P.H.'s performance is correct—an explanation in terms of the intactness of several stages of processing that normally precede conscious awareness of recognition. Then there are cognitive processes involved in face recognition, and Malcolm and Hamlyn seem to have committed themselves to a false empirical claim. What is more, they seem to have based the empirical claim upon the most flimsy evidence.

However, it would be uncharitable to suppose that these philosophers have made a gratuitous empirical claim by overreaching the proper domain of philosophy. Indeed, a philosopher of the cast of mind of Hamlyn or Malcolm would not accept that diagnosis of the situation, for he would not see himself as making an empirical claim at all. These philosophers are not engaged in pre-scientific empirical speculations. As we noted, Malcolm appears to be making a claim about what is constitutive of remembering; and, while the precise structure of the claim is not transparent, it is clearly meant to have an *a priori* character.

4.2 Three Responses

Given this construal of Malcolm and Hamlyn (as adherents of the second view of philosophy distinguished in the Introduction), we can imagine three

possible lines of response that these philosophers might offer to the case of P.H.

4.2.1 First Response: Failure of Imagination

A first line of response would be to accept that the claim that mental processes are a myth involves a philosophical error, and to regard the actual example of P.H. as making good a failure of philosophical imagination.

Given a false *a priori* claim, a counterexample is in principle constructible by *a priori* means; but there is no reason to expect any philosopher to be imaginatively equipped to generate every relevant kind of counterexample. As we noted in the Introduction, science may sometimes make good the imaginative shortfall. However, although the example of P.H. provides a vivid case in the actual world, no special weight attaches to its empirical character; there is still a fundamentally philosophical judgement to be made. On this line, it might be said that the situation with regard to P.H. is no different from what the situation would be if one of the science fiction examples so widely used in discussions of personal identity were to become scientific fact. Real world examples do not wear their interpretations on their sleeves; philosophical work is still required.

4.2.2 Second Response: Conceptual Confusion

Although this is a possible line of response, we do not suppose that our *a prioristic* philosophers of psychology would capitulate so rapidly. A second—and less concessive—line of response would be to say that cognitive science shows itself to be in the grip of a conceptual confusion by its claim that the patient P.H. is unconsciously or covertly recognizing faces. Hamlyn himself says that the proper domain of philosophy is ([1983], p. 304) ‘to provide a more meticulous description of different mental phenomena and work out what is implied in the concepts presupposed in our categorization of such phenomena’. And it is certainly open to someone to argue that some incoherence is involved in categorizing any phenomenon as both unconscious and a case of recognition.

However, it is no easy thing to make such an argument persuasive. It has to be said, surely, that the cognitive neuropsychological description of P.H., and the putative explanation of his particular constellation of symptoms, do not appear to be muddled or conceptually confused. Indeed, the friend of cognitive neuropsychology might turn the tables on this line of response by saying that it is precisely these careful case studies of cognitive deficits that can contribute to the ‘more meticulous description of different mental phenomena’.

If there is nothing especially confused about cognitive neuropsychology—nothing distinctively muddle-headed about the study of cognitive deficits—then the way to play out this second line of response is to argue for a quite general incoherence in the notion of unconscious psychological processes.

(This is, in effect, to adopt the third view that we distinguished in the Introduction.) Recently, John Searle has argued in just this way ([1989], p. 207):

Now oddly enough, this connection between consciousness and intentionality is lost in discussions of the unconscious mental processes in contemporary linguistics, philosophy, and cognitive science. Many of the phenomena which are cited as explanatory psychological features simply could not have any *psychological* reality because they are not the sort of things that could be mental states.

This is not the place for a detailed assessment of Searle's argument. Suffice it to remark that it is singularly difficult to establish in this way the nonexistence of the disciplines of linguistics and cognitive psychology. (See Searle [1990] with Peer Commentary.)

3.2.3 Third Response: Changing the Subject

A third line of response is to say that the psychologist is in some way changing the subject.

This is apparently the style of response that Malcolm [1959] offers to empirical work on dreaming and rapid eye movements. According to Malcolm, for 'the primary concept' of dreaming, 'the sole criterion of the occurrence of a dream is the waking report' ([1959], p. 70). That familiar concept makes no provision, for example, for the notion of the duration of a dream in physical time: so if psychologists introduce the notion of duration via investigation of rapid eye movements then what they have done is 'to create a new concept under an old label' (*ibid.*, p. 79). Indeed (*ibid.*, pp. 81–2):

Considering the radical conceptual changes that the adoption of a psychological criterion would entail, it is evident that a new concept would have been created that only remotely resembled the old one . . .

The desire to know more about dreaming should not lead scientists into transforming the concept in such a way that their subsequent discoveries do not pertain to *dreaming*.

The claim that scientific psychology changes the subject is certainly explicit here. But that claim is a consequence of the presumption of a very close constitutive tie between concepts and the criteria for their application. Once we reject the presumed theory of the individuation of concepts, Malcolm's version of the 'changing the subject' response comes to seem artefactual.

It does not follow, however, that the third line of response must lead rapidly to a dead end. There are several expressions of the 'changing the subject' claim—free, now, of any suggestion of a criteriological theory of meaning—in Hamlyn's work:

[I]t is no good insisting on the sanctity of scientific psychology . . . if it does not constitute an adequate psychological theory . . . (1990, p. 33)

[T]o get a proper view of what is required [for so-called cognitive science] we need to take account of the whole complex of input, intervening processes and output . . . If it is claimed that we do not need to note what is implied in the notion of behaviour, as we ordinarily understand that, an alternative construal of output must be forthcoming. Any alternative expressed in terms appropriate to physics and physiology alone gives rise to all the objections on the score of overwhelming complexity noted earlier, apart from the fact that such an account would have *no connection with folk psychology and everything that would make the account psychologically relevant.* (*ibid.*, p. 37; our emphasis)

In order to avoid caricaturing the practice of scientific psychology, we need to be careful to distinguish cognitive psychological descriptions from physiological descriptions. But the suggested shape of the third line of response is clear enough.

Recall that the explanation of P.H.'s performance on the name categorization task appeals to a degree of distraction by the face of a person whose occupation is different from that of the person named. This is an explanation cast in psychological terms rather than physiological terms. The explanation involves the idea of accessing information about some of the properties of the person whose face is shown.

An advocate of the third line of response could agree that there are indeed such processes as are invoked in that explanation, and that they do indeed require description at some level other than physiological; but he would still insist that these processes must not be described as *mental*.

However, this appears, at first sight, to be a merely stipulative, terminological manoeuvre. It seems to rob the third line of response of all theoretical interest; and it could hardly substantiate the claim that the idea of mental processes is a myth.

4.3 Two Explanatory Styles

In order to defend the theoretical importance of the terminological recommendation—that the term 'mental' should not be extended to unconscious or covert information processing—it would be necessary to elaborate further the claim that the psychologist is changing the subject.

It clearly will not do to say that the psychologist offers no account at all of our mental life, but only an account of our brain processes—an account that is so far removed from the domain of folk psychology as to be irrelevant to our philosophical concerns. That is simply to beg the question against the existence of a distinctive discipline of *cognitive* psychology. But let us, for a moment, sympathetically explore the idea that there are theoretically important differences between the personal level, folk psychological domain of mentation and the subpersonal level, cognitive psychological domain of information processing.

John McDowell has stated explicitly that it is not clear that an account in

information processing terms has anything to do with the *mind* at all. This judgement stems from a particular conception of the folk psychological arena, and especially of the explanation of intentional action. For McDowell, like Donald Davidson, stresses that such explanations rationalize action.

the concepts of the propositional attitudes have their proper home in explanations of a special sort: explanations in which things are made intelligible by being revealed to be, or approximate to being, as they rationally ought to be. This is to be contrasted with a style of explanation in which we make things intelligible by representing their coming into being as a particular instance of how things in general tend to happen. (McDowell [1985], p. 389)

On this view, what could be wrong with the putative information processing explanation of the prosopagnosia exhibited by P.H.?

McDowell might express his concerns in this way. The attempt, in subpersonal models of psychological processing, to drill below bedrock (to use Wittgenstein's evocative phrase) seeking strata that are free from normativity, is doomed to failure. For the drilling merely allows the intentional notions used at the personal level—and hence also their normative accompaniments—to trickle down to the subpersonal strata. Therefore, any account of the subpersonal type will be parasitic upon the intentional language whose proper home is at the personal level.

McDowell can be construed as arguing that in subpersonal psychology we invert the computational metaphor. Thus, the cognitive psychologist is not explaining the personal by the subpersonal; rather, she is explaining the subpersonal via the personal. It is not that the mind is analogous to a computer program, but that a computer program is analogous to the mind. McDowell himself locates the target of his remarks about rationalization as follows ([1985], p. 397):

[They] are only critical of functionalism as a theory of propositional attitudes: they do not tell against functionalism as a framework for a theory about how sub-personal states and events operate in the control of behaviour, although they do raise a question about what (if anything) that sort of theory has to do with the mind.

There is a great deal here that calls for discussion. On the one hand, we could agree that a description of the causal role of states in the production of behaviour will not yield a reductive account of our mental life with all its normative aspects. On the other hand, if McDowell is suggesting that the only way that subpersonal psychology can make use of notions of semantic content is by invoking little men and women (*homunculi*) performing intentional actions—including acts of communication—then, to say the least, his case is not made out.

Perhaps, in the present context, we can leave some of these issues unresolved. The pivotal point in McDowell's argument is surely the claim that

'the concepts of the propositional attitudes have their proper home in explanations of a special sort: explanations in which things are made intelligible by being revealed to be, or approximate to being, as they rationally ought to be'. For this claim suggests a principle of distinction between the folk psychological and the cognitive psychological domains.

As we have already mentioned, McDowell traces the claim about a distinctive style of rationalizing explanation back to Davidson's work in the theory of radical interpretation. But very similar claims have been made by philosophers of many different schools, who are united solely by their desire to drive a wedge between the kind of explanation appropriate in the natural sciences and that appropriate in the social and human sciences (see, for example, Winch [1958]; Taylor [1985]).

Although this is to simplify complex discussions, we can say that, in all of these types of account, explanation in the field of the human or social sciences is thought to demand that the investigator empathize with the object of investigation. This general strategy is elucidated by Jane Heal, when she says ([1986], p. 137): 'I can harness all my complex theoretical knowledge about the world and my ability to imagine to yield an insight into other people without any further elaborate theorising about them.' The major challenge to philosophers who wish to make out this case has been to give a convincing account of what this empathetic relationship consists in. Davidson's [1973, 1974] theory of radical interpretation might be thought of as providing just such an account.

In brief, Davidson argues that to understand someone else's behaviour (linguistic or otherwise) requires the radical interpreter to triangulate that person's beliefs and desires, his intentions, and the meaning of his utterances. I can come to an understanding of someone's beliefs if I can understand the meaning of his utterances; and I can understand what someone says if I know what belief prompts the utterance—all this, of course, against the background of an understanding of the person's intentions. The problem is to find a fixed point from which the triangulation can be carried out. Davidson argues that the only such point is constituted by the beliefs and desires of the interpreter herself. He makes this beautifully clear when he says ([1982], p. 302):

[T]he strategy [of radical interpretation] is to assume that the person to be understood is much like ourselves. That is perforce the opening strategy, from which we deviate as evidence piles up. We start out assuming that others have, in the basic and largest matters, beliefs and values similar to ours. We are bound to suppose someone we want to understand inhabits our world of macroscopic, more or less enduring, physical objects with familiar causal dispositions; that his world, like ours, contains people with minds and motives; and that he shares with us the desire to find warmth, love, security, and success, and the desire to avoid pain and distress.

So, if the interpreter is allowed to assume that the interpretee is sufficiently like

her in various respects, then when he is faced with what she would recognize as a rabbit, she can surmise (at least as an initial hypothesis) that he sees the world as containing a rabbit—that he believes that there is a rabbit. In effect, what the radical interpreter is doing here is empathizing—saying to herself: if I were myself in the position of the interpretee, then what would I come to believe?

According to the Davidsonian scheme, this theory of interpretation is constitutive of mental life, of action, and of what it is to make a meaningful utterance. If a being's vocal blasts are not interpretable via the method of radical interpretation, then they are not speech; if intelligible attitudes cannot be attributed to a being via that method, then the being is not minded.

If these broad brush strokes capture the general thrust of the Davidsonian position, and if this is what lies behind McDowell's remarks about the irrelevance of cognitive psychology to the mental domain, then what shall we say in reply? We offer two brief comments.

First, it is clear that the cognitive neuropsychological explanations of the behaviour of P.H. are not grounded in the methodology of radical interpretation. There are two reasons for this. On the one hand, cognitive neuropsychological accounts depend upon the development of models of normal cognitive processing (see Section 2.1 above). As a straightforward matter of fact, such models—for example, Marr's [1982] theory of vision, or the dual route model of reading words aloud—are not developed via the method of radical interpretation. On the other hand, cognitive deficits simply do not lend themselves to the methodology of radical interpretation. Indeed, it is partly their interpretative mysteriousness that motivates the cognitive neuropsychological project.

Second, disorders of cognition may oblige us to incorporate into a folk psychological description elements drawn from the scientific psychological domain. An example is provided by Davidson himself [1982]. Suppose we try to provide an account of certain kinds of irrationality—wishful thinking, self-deception—within a general interpretative framework that attaches great importance to the constitutive role of norms of rationality: a framework in which explanations of mental events under their mental descriptions are always rationalizing explanations. Then it is more or less inevitable that we should import a partitioning of the mind—a simple form of the modularity hypothesis.

Mark Johnston [1988] rejects Davidson's account of these cases of irrationality, and replaces Davidson's appeal to modularity—or homuncularism—with a different departure from the 'interpretive view [which] counts rationality as both constitutive and exhaustive of the mental' ([1988], p. 80). Johnston's claim is that we have to recognize non-rationalizing patterns of causation amongst mental states. We should regard wishful thinking, for example (*ibid.*, p. 73): 'as a mental mechanism . . . by which a desire that *p* and

accompanying anxiety that not-*p* set the conditions for the rewarding (because anxiety-reducing) response of coming to believe that *p*.' Thus the 'style of explanation in which we make things intelligible by representing their coming into being as a particular instance of how things in general tend to happen' (McDowell [1985], p. 389) intrudes into the folk psychological arena.

The upshot is this. We can agree that there is a style of explanation that is characteristic of personal level folk psychology, and different from the explanation by subsumption that is typical of subpersonal level cognitive psychology—as of other sciences. Someone might base a terminological recommendation upon this difference: that the term 'mental' should be used to mark out the personal level folk psychological domain. But one can hardly use this difference of explanatory styles to ground the claim that subpersonal cognitive psychology is irrelevant to personal level folk psychology. For the style of explanation that is typical of the scientific domain also has to be employed in the folk domain. In short, a terminological distinction between the *genuinely mental* domain of folk psychology and the subpersonal domain of information processing, if it is based upon difference of explanatory style, does nothing to suggest the constitutive irrelevance of one domain to the other.

We have been examining challenges to the idea that there are mental processes implicated, for example, in face recognition: philosophical challenges which would—in the limit—consign to mythology the cognitive psychological accounts of face recognition in normal subjects, and of P.H.'s cognitive deficit. Philosophers mounting such a challenge might *per improbabile* simply concede when presented with the case study of P.H. (4.2.1), or they might attempt a global attack on the very coherence of the discipline of cognitive psychology (4.2.2). Alternatively, they might accept that cognitive psychology is all very fine in its way, but insist that its way simply bypasses the topic of real mentation (4.2.3).

If this insistence is to be justified, then some argument must be given to show that subpersonal level cognitive psychology is constitutively irrelevant to personal level folk psychology. We have just seen that it is impossible to argue that the two pass each other by, simply on the basis of a distinction between explanatory styles. But, of course, that impossibility does not settle the question of the relationship between the two domains. It remains to consider, more generally, whether empirical discoveries about subpersonal information processing can impinge upon our folk and philosophical conceptions of ourselves as persons.

4.4 *Independence or Interaction*

There certainly are differences between personal level folk psychology and subpersonal level scientific psychology. It is no part of our aim to minimize these differences; still less to suggest that there can be a satisfying reductive

account of every aspect of folk psychology cast in information processing terms. We do, however, wish to confront head-on the question whether these acknowledged differences render the second domain constitutively irrelevant to the first.

Recall the name categorization task (3.3): subjects are asked to categorize names by occupation, while pictures of faces are also presented. Normal subjects respond more slowly when the face is of someone with a different occupation from the person named than when the face is of someone with the same occupation. Consider what might be the ordinary folk psychological explanation of the interference effect in the case of a normal subject.

We must first acknowledge that taking longer to respond in one condition than another is not, of course, something that the subject does intentionally. Indeed, normal subjects take longer to respond in the *Unrelated* condition even though they are asked to ignore the faces. So the interference effect does not have a strictly rationalizing explanation in terms of reasons for action.

But, nevertheless, someone who adopts the stance of the radical interpreter may very well say that the interference effects is *intelligible* given normal expectations of rationality: the concept of distraction is perfectly at home in the folk domain. If a normal subject recognizes the face presented and knows that it is the face of a politician, then it is no surprise that this knowledge should interfere with his ability to categorize a simultaneously presented name as that of a TV personality. If the interpreter imagines herself in the place of the subject, then she can empathetically understand well enough how the presentation of the face of Neil Kinnock could impede categorization of 'Terry Wogan' as the name of a TV personality.

For P.H. as for normal subjects, there is a marked difference between the condition in which the presented face is of someone whose occupation is different from the person named, and the condition where the face and the name are of people with the same occupation. (There is a significant difference in reaction times between the *Unrelated* and *Name only* conditions, but not between the *Related* and *Name only* conditions; see again Table 2.) P.H. shares this interference effect in the *Unrelated* condition with normal subjects; but the folk psychological style of explanation is wholly inadequate for P.H. In his case, the interference effect occurs without any conscious recognition of the face as the face of Neil Kinnock, or even as the face of a politician. (Recall that P.H. performs virtually at chance levels when asked to categorize the faces.) Here, rationalizing interpretation and empathetic imagination draw a blank.

So, although there are notable similarities between P.H.'s performance and that of normal subjects on the name categorization task (as on the face matching task) there is no prospect of a folk psychological explanation of the interference effect that generalizes across P.H. and normal subjects.

It seems to be a coherent hypothesis—coherent, though not, of course *a priori* true—that the explanation of these effects in P.H. is just the same as in

normal subjects. Suppose for a moment that this empirical hypothesis is correct. Then—we have just seen—the common explanation cannot be folk psychological. But nor can it be neurophysiological. For, first, there is no guarantee that there are always physiological commonalities matching similarities in cognitive performance. And, second, a physiological explanation cannot explain the cognitive effects *as such*; that really would be a case of changing the subject.

So—continuing with the presumably coherent supposition that the interference effect has just the same explanation in normal subjects and in P.H.—let us now suppose that the common explanation is a cognitive psychological explanation. Then the question that presses is what the relation can be between the *two* explanations that both apply to a normal subject. There is a folk psychological explanation in terms of overt recognition; and there is a cognitive psychological explanation in terms of levels of unconscious information processing that are amongst the causal antecedents of overt recognition.

The constitutive irrelevance of subpersonal scientific psychology to folk psychology requires that these two explanations should simply pass each other by. But, on the face of it, they are not logically independent of each other. Since folk psychological explanations are still a kind of causal explanation, it is part of the folk psychological explanation that the interference effect is causally downstream of the conscious recognition of Neil Kinnock's face as that of a politician. But the whole burden of any explanation that can also subsume the case of P.H. is that the interference effect cannot be causally downstream of conscious recognition. For with P.H. interference takes place, while conscious recognition does not. To that extent, the two explanations that apply to a normal subject are in competition.

Thus, a coherent empirical hypothesis can bring an explanation in terms of subpersonal information processing into competition with a natural folk psychological explanation. The folk psychological explanation might, in principle, be shown to be incorrect by empirical discoveries in cognitive psychology. This is enough to rebut the global claim that subpersonal level scientific psychology is logically independent of, and constitutively irrelevant to, personal level folk psychology.

Our claim that there is interaction, rather than independence, between folk psychology and scientific psychology is a modest one; and it is as well to enter three clarificatory comments.

First, it is no part of our claim that the explanation of the interference effect in normal subjects is in fact the same as the explanation of the interference effect in P.H. We are saying merely that, if the explanation is just the same in the two cases, then that shared explanation competes with the folk psychological explanation in terms of overt recognition and distraction in the case of a normal subject.

The possibility that the explanations are in competition is all we need to

make our point. So far as we know, it is just as plausible that there is actually no competition. There might be two components to the interference effect in normal subjects—one component from the conscious experience of recognition and another from unconscious information processing—only one of which is present in P.H. In that case (as Ned Block has pointed out to us) the difference between the circumstances of a normal subject and the circumstances of P.H. would be analogous to the difference between supraliminal and subliminal advertising.

Second, it is worth noting that, on that non-competitive option, the folk psychological explanation does not provide a complete explanatory story about the interference effect in normal subjects. (It might seem that there is another non-competitive option, upon which the folk psychological explanation is the whole story about the interference effect in normal subjects, while the correct explanation of the effect in P.H. is in terms of unconscious processes. But (as Max Coltheart has pointed out to us) this option conflicts with the transparency assumption, introduced earlier (2.1). For it would have to be that brain damage had introduced a new effect in P.H.—an effect that is absent from the intact face recognition system in normal subjects.)

This possible explanatory incompleteness of folk psychology reinforces the point that we took from Johnston [1988]; namely, that we must recognize explanations of the subsuming rather than the rationalizing style, even within the mental arena. A further example is close to hand. We said that someone adopting the stance of the radical interpreter could find the interference effect in normal subjects intelligible. Faces interfere with the categorization of names; but it is also the case that names interfere very much less with the categorization of faces (Young *et al.* [1986], p. 469). It is far from easy to see how to elaborate the folk psychological explanation to account for this asymmetry.

Third, it is no part of our claim that, if the coherent empirical hypothesis is in fact correct, then the whole folk psychological scheme is endangered. On the contrary, however natural the folk psychological explanation of the interference effect in a normal subject may be, it can scarcely be reckoned as a central or essential component of the folk scheme. Indeed, it is not immediately obvious how much that is essential to the folk scheme would be endangered if it were to be discovered that, in many cases, what we take to be causal consequences of conscious states are really causal consequences of information processing states that are dissociable from conscious awareness. The discovery that conscious states are epiphenomenal would surely disturb some of our common-sense views about the way the world works; but it is far from clear that such a discovery would license any dramatically eliminativist conclusions.

With these three comments, it should be clear that the example that we have sketched lies at a point near one end of a spectrum of cases. It is an example in

which an empirical discovery in cognitive psychology could come into conflict with a component—neither central nor essential—of the folk psychological scheme. Near the other end of the spectrum would be cases in which cognitive psychological discoveries would come into conflict with components of the folk scheme that are revealed by our best available philosophy of mind to be intrinsic to our very notion of a thinking subject.

The points on this spectrum all instantiate a general scheme. Our folk psychological descriptions incur causal commitments. Generally subpersonal psychology reveals how those commitments are met (Peacocke [1992], Chapter 7); but, on occasion, cognitive psychology may show that the causal commitments of our folk descriptions are not met (Ramsey, Stich, and Garon [1990]; Davies [1991]). In that case, our folk conception of ourselves needs to undergo revision.

That concludes our main line of argument in this paper. Having set out some of the theoretical commitments of cognitive neuropsychology in Section 2, and summarized the case study of P.H. in Section 3, we have in this section been arguing for the relevance of the science of cognitive psychology to the understanding of our mental lives.

The cognitive psychological model of face recognition that is supported by the case of P.H. illustrates in a small way that it is impossible to sustain the claim that personal level folk psychology and subpersonal level scientific psychology are independent of, and irrelevant to, each other. For not only does that model enable us to capture a generalization about cognitive performance (for example, on the name categorization task) that is beyond the scope of the folk psychological scheme; the model also shows us how folk psychological explanations of the performance of normal subjects might be wrong.

In the end, this also motivates a particular conception of the relationship between philosophy and psychology. We regard these two components of contemporary cognitive science as disciplines with quite distinctive methodologies, standing in an interactive relationship. Often, philosophical theory uncovers necessary conditions for the application of personal level folk psychological properties. Subpersonal level cognitive psychology tells us whether, and if so how, those necessary conditions are met.

5 PHILOSOPHICAL LESSONS FROM COGNITIVE NEUROPSYCHOLOGY

In this final section, we briefly suggest four further examples of the philosophical significance of cognitive neuropsychology.

5.1 *The Modularity of Mind*

One of the theoretical assumptions underlying research in cognitive neuropsychology is a thoroughgoing modularity. The success of the enterprise is consequently a vindication of that assumption.

The claim that the mind is modular in structure cannot be justified on *a priori* grounds alone; it is an empirical hypothesis. But we would claim that it is the type of large-scale claim that cannot plausibly be denied philosophical interest.

5.2 *The Refinement of Mind*

Cognitive neuropsychology provides us with a much more fine-grained account of the mental than that given us by those pre-theoretical intuitions upon which many of our philosophical claims are (inevitably) based. There is a tendency in philosophy to think in terms of large, poorly differentiated, categories. Indeed, the monolithic notion of the mental or the mind is itself an example. The multiple dissociations revealed in cognitive neuropsychological case studies make us realize that the phenomenon of mind is much more complex and heterogeneous than we may initially be disposed to believe.

5.3 *The Elimination of Mind*

We have argued that the cognitive neuropsychological demonstration of the existence of unconscious recognition reveals an interesting and important limitation on the scope of our folk psychological scheme.

We do not make the sweeping claim that cognitive psychological explanations should replace folk psychological explanations across the board. The situation is rather more complex than that.

Let us agree that our folk psychological scheme serves our predictive purposes well. Nevertheless, it may be the concepts used there carry causal and ontological commitments which science reveals not to be met. This situation need not be regarded as showing that a philosophical error has been committed. The philosopher's account of what is involved in our folk psychological conception of the mental domain may be wholly accurate. But it may just be that there is nothing in reality that precisely meets the conditions that the conceptual analysis articulates.

5.4 *Consciousness*

Our final suggestion is that work in cognitive neuropsychology provides resources for some progress with a topic of undoubted philosophical interest—indeed, a philosophical topic that must be considered an archetype—namely, consciousness—its nature and purpose.

There is no remotely satisfactory philosophical account of the nature and functions of consciousness. So the merest suggestion of ways in which cognitive neuropsychological findings may be of help serves to strengthen the case for interdisciplinary research.

There is a quite natural thought about consciousness which we take to be undermined by the work we have presented, and much other work in a similar vein. This is the thought that consciousness is a global and uniform

phenomenon. We note that even a philosopher who is committed to the 'cognitive revolution'—indeed perhaps synonymous with it: we mean Jerry Fodor—appears to take consciousness as a global affair.

In his essay *The Modularity of Mind* [1983] Fodor articulates and defends a modular account of the mind that is, in general terms, not unrelated to the account that we are drawn to on the basis of the neuropsychological work. However, Fodor argues that the modular organization of the mind is not total. He makes a distinction between input systems that are modular, and the non-modular, holistic, central system.

The input systems perform relatively low-level processing tasks—*e.g.* processing an acoustic blast into a logical form, or a pattern of retinal irradiation into a two-and-a-half-D sketch. However, the judgement that an utterance of 'He's brave isn't he?', means that he's a coward, or that the scene before one's eyes means that there's a leopard in the vicinity, are made by the non-modular central system on the basis of the evidence provided by the input systems calibrated with background knowledge.

A tempting image has the various input systems delivering their respective outputs (such as logical forms or two-and-a-half-D sketches) to a homogeneous central system that is uniformly floodlit by consciousness. But the *existence of highly selective deficits of conscious awareness forces us to think more carefully about that image.*

We do not say that the fact that overt face recognition can be abolished whilst other conscious processes remain (P.H. remains aware that the name 'Neil Kinnock' is the name of a politician) by itself puts pressure on the view that consciousness is global and uniform. For it may be possible to account for this dissociation by saying that, in P.H., the face recognition input system has become disconnected from the central system, and is no longer able to pass its outputs into the floodlit arena (Schacter *et al.* [1988]). But still, the wide variety of failures of awareness at different levels of visual processing do put pressure on the view that consciousness is global and uniform (Young and De Haan [1990], pp. 43–4). For example, the simple idea of disconnection between an input system and the central cognitive system would not easily account for a case in which both earlier and later stages of processing within an input system were intact, but where only the earlier stages were available to consciousness. De Haan *et al.* [1992] report on a prosopagnosic patient N.R. who may fit this description. N.R. differs from P.H. in one striking way. N.R. has a severe impairment of overt face recognition, accompanied by evidence of at least some degree of covert recognition. But, unlike P.H., his performance is well above chance when he is asked to choose which of two faces (one familiar, one unfamiliar) is the familiar one. Thus, N.R. shows some measure of awareness of the familiarity of a face, even though he is quite unaware whose face it is.

The results of cognitive neuropsychological case studies also undermine a

second natural thought about consciousness; namely, the thought that conscious mental processes must be underpinned by, or justified in terms of, further conscious mental processes. In an influential paper, Anthony Marcel [1983] argues, on the basis of a large amount of data from the study of normal subjects and data similar to that used here, that what he calls the *Identity Assumption* is false. The Identity Assumption is the claim that ([1983], p. 238) 'the representations which constitute conscious experience are . . . the very same ones that are derived and used in sensory and cognitive processing.' We do not want to go into the details of Marcel's rich discussion; it is clear enough how case studies like that of P.H. tell against the Identity Assumption. It is also clear that, once the Identity Assumption is rejected, it is much easier to separate questions about subpersonal level cognitive processing leading up to a conscious mental state from questions about personal level conscious antecedents of that conscious mental state. Consequently, we find it appealing—though undoubtedly anachronistic—to see in the Identity Assumption an encapsulation of the undertow against which Wittgenstein himself was fighting; an undertow by which neo-Wittgensteinian philosophers like Malcolm are—unbeknown to themselves—being swept along.

6 CONCLUSION

In this paper we have tried to persuade you of the philosophical interest of cognitive neuropsychology.

There are two ways in which it is of interest. First, and surely uncontroversially, philosophers can take a metatheoretical interest in the assumptions and arguments of cognitive neuropsychology (Section 2.3).

However, since we do not accept the view that philosophy is exclusively metatheoretical, our main concern has been with a second way in which cognitive neuropsychology has interest for philosophers. We argued that it provides insight into issues that are clear cases of philosophical problems.

In Section 4, we argued that some influential claims in the philosophy of mind—allegedly stemming from Wittgenstein's later work—are shown to be untenable. In Section 5, we listed and briefly commented upon some other topics of general philosophical interest for which cognitive neuropsychology promises to have consequences.¹

*King Alfred's College
Winchester SO22 4NR
Corpus Christi College
Oxford OX1 4JF*

REFERENCES

- BAKER, G. P. and HACKER, P. M. S. [1984]: *Language, Sense and Nonsense*. Oxford: Basil Blackwell.
- BERLIN, I. [1980]: *Concepts and Categories*. Oxford: Oxford University Press.
- BUB, J. and BUB, D. [1988]: 'On the Methodology of Single-case Studies in Cognitive Neuropsychology', *Cognitive Neuropsychology*, 5, pp. 565–82.
- BUB, D., CANCELLIERE, A. and KERTESZ, A. [1985]: 'Whole-word and Analytic Translation of Spelling to Sound in a Non-semantic Reader', in K. E. Patterson, J. C. Marshall and M. Coltheart (eds.): *Surface Dyslexia: Neurophysiological and Cognitive Studies of Phonological Reading*. London: Lawrence Erlbaum Associates.
- CARAMAZZA, A. [1984]: 'The Logic of Neuropsychological Research and the Problem of Patient Classification of Aphasia', *Brain and Language*, 21, pp. 9–20.
- CARAMAZZA, A. [1986]: 'On Drawing Inferences About the Structure of Normal Cognitive Systems from the Analysis of Patterns of Impaired Performance: A Case for Single-Patient Studies', *Brain and Cognition*, 5, pp. 41–66.
- COLTHEART, M. [1985]: 'Cognitive Neuropsychology and the Study of Reading', in M. I. Posner and G. S. M. Martin (eds.): *Attention and Performance XI*. Hillsdale, NJ.: Lawrence Erlbaum Associates.
- COLTHEART, M., SARTORI, G. and JOB, R. [1987]. *The Cognitive Neuropsychology of Language*. Hillsdale, NJ.: Lawrence Erlbaum Associates.
- CUMMINS, R. [1989]: *Meaning and Mental Representation*. Cambridge, MA.: MIT Press.
- DAVIDSON, D. [1973]: 'Radical Interpretation'. in *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press [1984].
- DAVIDSON, D. [1974]: 'Belief and the Basis of Meaning', in *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press [1984].
- DAVIDSON, D. [1982]: 'Paradoxes of Irrationality', in R. Wollheim and J. Hopkins (eds.): *Philosophical Essays on Freud*. Cambridge: Cambridge University Press.
- DAVIES, M. [1991]: 'Concepts, Connectionism, and the Language of Thought', in W. Ramsey, S. P. Stich and D. E. Rumelhart (eds.): *Philosophy and Connectionist Theory*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

¹ David Hamlyn (Professor of Philosophy at Birkbeck College until 1988) provided the initiative for interdisciplinary liaisons at Birkbeck, and demonstrated in his own work the value of philosophical engagement with empirical work in psychology. Max Coltheart (Professor of Psychology at Birkbeck College until 1987) taught us most of what we know about cognitive neuropsychology. We are deeply grateful to them both.

An early version of this paper was presented to a meeting of the British Society for the Philosophy of Science, 13 November 1989. Subsequently, talks based on the material have been given at the Technological University of Compiègne, the University of Maryland, the University of New South Wales, and the University of Oregon. Thanks to members of the audiences in those places for their comments.

Andy Young provided us with detailed comments on a penultimate draft, saving us from a number of errors and leading to other improvements. We are grateful for his guidance.

The connection between interpretation and empathy, which is briefly mentioned in Section 4.3, suggests a way of developing the issue of a different explanatory style which is the subject of an expansive debate in both philosophy and psychology. See the special issue of *Mind and Language* on Mental Simulation (Volume 7, Numbers 1 and 2, 1992), and our forthcoming collection *Mental Simulation: Philosophical and Psychological Essays* (Oxford: Blackwell Publishers).

- DE HAAN, E. H. F., YOUNG, A. W. and NEWCOMBE, F. [1987]: 'Face Recognition Without Awareness', *Cognitive Neuropsychology*, **4**, pp. 385–415.
- DE HAAN, E. H. F., YOUNG, A. W. and NEWCOMBE, F. [1992]: 'Neuropsychological Impairment of Face Recognition Units', *Quarterly Journal of Experimental Psychology*, **44A**, pp. 141–75.
- ELLIS, A. W. [1987]: 'Intimations of Modularity, or, the Modularity of Mind: Doing Cognitive Neuropsychology Without Syndromes', in Coltheart, Job and Sartori [1987].
- ELLIS, A. W. and YOUNG, A. W. [1988]: *Human Cognitive Neuropsychology*. Hove, East Sussex: Lawrence Erlbaum Associates.
- FODOR, J. [1983]: *The Modularity of Mind*. Cambridge, MA.: MIT Press.
- FUNNELL, E. [1983]: 'Phonological Processes in Reading: New Evidence from Acquired Dyslexia', *British Journal of Psychology*, **74**, pp. 159–80.
- HAMLIN, D. W. [1983]: *Perception, Learning and the Self*. London: Routledge & Kegan Paul.
- HAMLIN, D. W. [1990]: *In and Out of the Black Box: On the Philosophy of Cognition*. Oxford: Basil Blackwell.
- HARRIS, M. and COLTHEART, M. [1986]: *Language Processing in Children and Adults*. London: Routledge & Kegan Paul.
- HEAL, J. [1986]: 'Replication and Functionalism', in J. Butterfield (ed.): *Language, Mind and Logic*. Cambridge: Cambridge University Press.
- HUMPHREYS, G. W. and BRUCE, V. [1989]: *Visual Cognition: Computational, Experimental, and Neuropsychological Perspectives*. Hove, East Sussex: Lawrence Erlbaum Associates.
- HUMPHREYS, G. W. and EVETT, L. J. [1985]: 'Are There Independent Lexical and Non-lexical Routes in Word Processing? An Evaluation of the Dual-route Theory of Reading', *Behavioral and Brain Sciences*, **8**, pp. 689–740.
- JOHNSTON, M. [1988]: 'Self-deception and the Nature of Mind, in B. P. McLaughlin and A. O. Rorty (eds.): *Perspectives on Self-deception*. Berkeley and Los Angeles: University of California Press.
- MCDOWELL, J. [1985]: 'Functionalism and Anomalous Monism', in E. LePore and B. McLaughlin (eds.): *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell.
- MALCOLM, N. [1959]: *Dreaming*. London: Routledge & Kegan Paul.
- MALCOLM, N. [1971]: 'The Myth of Cognitive Processes and Structures', in *Thought and Knowledge*. Ithaca: Cornell University Press [1977].
- MARCEL, A. J. [1983]: 'Conscious and Unconscious Perception: An Approach to the Relations between Phenomenal Experience and Perceptual Processes', *Cognitive Psychology*, **15**, 238–300.
- MARR, D. [1982]: *Vision*. New York: W. H. Freeman & Co.
- PEACOCKE, C. [1992]: *A Study of Concepts*. Cambridge, MA.: MIT Press.
- PETERSEN, S. E., FOX, P. T., POSNER, M. I., MINTUN, M. and RAICHLER, M. E. [1988]: 'Positron Emission Tomographic Studies of the Cortical Anatomy of Single-word Processing', *Nature*, **331**, pp. 585–9.
- PETERSEN, S. E., FOX, P. T., SNYDER, A. Z. and RAICHLER, M. E. [1990]: 'Activation of Extrastriate and Frontal Cortical Areas by Visual Words and Word-like Stimuli', *Science*, **249**, pp. 1041–4.

- RAMSEY, W., STICH, S. and GARON, J. [1990]: 'Connectionism, Eliminativism and the Future of Folk Psychology', in J. E. Tomberlin (ed.): *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*. Atascadero, C.A.: Ridgeview Publishing Company.
- SCHACTER, D. L., MCANDREWS, M. P. and MOSCOVITCH, M. [1988]: 'Access to Consciousness: Dissociations between Implicit and Explicit Knowledge in Neuropsychological Syndromes', in L. Weiskrantz (ed.), *Thought Without Language*. Oxford: Oxford University Press.
- SEARLE, J. R. [1989]: 'Consciousness, Unconsciousness, and Intentionality', *Philosophical Topics*, 17, pp. 193–209.
- SEARLE, J. R. [1990]: 'Consciousness, Explanatory Inversion and Cognitive Science' (with Peer Commentary and Author's Response), *Behavioral and Brain Sciences*, 13, pp. 585–642.
- SHALLICE, T. [1988]: *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.
- SHALLICE, T. and SAFFRAN, E. [1986]: 'Lexical Processing in the Absence of Explicit Word Identification: Evidence from a Letter-by-letter Reader', *Cognitive Neuropsychology*, 3, pp. 429–58.
- TAYLOR, C. [1985]: *Philosophical Papers Volume 2: Philosophy and the Human Sciences*. Cambridge: Cambridge University Press.
- WINCH, P. [1985]: *The Idea of a Social Science*. London: Routledge & Kegan Paul.
- YOUNG, A. W. [In press]: 'Covert Recognition', in M. Farah and G. Ratcliff (eds.): *The Neuropsychology of Higher Vision: Collected Tutorial Essays*. Hillsdale, NJ.: Lawrence Erlbaum.
- YOUNG, A. W. and DE HAAN, E. H. F. [1988]: 'Boundaries of Covert Recognition in Prosopagnosia', *Cognitive Neuropsychology*, 5, pp. 317–36.
- YOUNG, A. W. and DE HAAN, E. H. F. [1990]: 'Impairments of Visual Awareness', *Mind and Language*, 5, pp. 29–48.
- YOUNG, A. W. and DE HAAN, E. H. F. [1991]: 'Face Recognition and Awareness After Brain Injury', in A. D. Milner and M. Rugg (eds.): *The Neuropsychology of Consciousness*. London: Academic Press.
- YOUNG, A. W., ELLIS, A. W., FLUDE, B. M., MCWEENEY, K. H. and HAY, D. C. [1986]: 'Face-Name Interference'. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 466–75.
- YOUNG, A. W., HELLAWEEL, D. and DE HAAN, E. H. F. [1988]: 'Cross-domain Semantic Priming in Normal Subjects and a Prosopagnosic Patient', *Quarterly Journal of Experimental Psychology*, 40A, pp. 561–80.
- YOUNG, A. W., MCWEENEY, K. H., HAY, D. C. and ELLIS, A. W. [1986]: 'Matching Familiar and Unfamiliar Faces on Identity and Expression', *Psychological Research*, 48, pp. 63–8.