

---

# ANALYTICAL BEHAVIOURISM

## Objectives

As a result of reading this chapter you should:

- understand the difference between psychological and philosophical or analytical behaviourism;
- understand the central claim of analytical behaviourism, namely that talk about mental states is equivalent in meaning to talk about actual and potential patterns of behaviour;
- understand the strengths of analytical behaviourism;
- understand Hempel's 'hard' version of behaviourism and the factors which motivated it;
- understand and appreciate the importance of avoiding circularity when attempting an analytical reduction of mental states;
- understand the difficulties Hempel faces in trying to provide a non-circular behavioural analysis of mental states;
- understand Ryle's 'soft' version of behaviourism and the factors which motivated it;
- understand Ryle's account of a disposition and be able to evaluate it;
- understand the difficulty behaviourism faces when trying to provide explanations of actions and behaviour in general;
- understand the difficulty behaviourism faces in accounting for the phenomenology of non-intentional states and its apparent privacy;
- understand why behaviourism cannot provide for first-person authority regarding the individual's knowledge of his own states of mind;
- understand why the possibility of pretence cannot show that behaviourism is false, but appreciate the strengths of the arguments from zombies and super-Spartans in achieving this goal.

## 4.1 Introduction

Dualism and the mind/brain identity theory share the assumption that the mind is a thing, a non-physical Cartesian substance, on the one hand, and the living brain, on the other. Analytical behaviourism challenges and rejects this presumption, in common with functionalism, which will be explored in the next chapter. Most straightforwardly and simply, analytical behaviourism maintains that statements about the mind and mental states turn out, after analysis, to be equivalent to statements that describe a person's actual and potential public behaviour. There is, on this view, ultimately no more to someone's mental states than certain overt patterns of behaviour he or she exhibits, or, in the appropriate circumstances, is disposed to manifest.<sup>1</sup>

Analytical behaviourism, if it can be made to work, has a number of strengths. These are:

*Avoidance of the mind/body interaction problem* Firstly, it makes the nature of the relationship of the mind to the body perspicuous: the mind just is the behaviour, actual and potential, of the body. The mind does not cause the behaviour: it *is* the behaviour. Worries about the nature of the mind conceived of as an immaterial substance cannot arise.

*The non-mysteriousness of the mental* There are no mysterious obscure mental properties to account for. Should a behaviourist be at all concerned about the relation of the mind to the brain, the problem reduces to the purely scientific one of which internal physical states are causally responsible for various patterns of behaviour. All speculation about how non-physical properties can emerge from underlying physical processes is rendered redundant.

*Dissolving the problem of other minds* Cartesian dualism would appear to make knowledge of mental states other than one's own impossible. Mental states comprise a radically private realm to which only the person whose states of mind they are has access. Behaviourism sidesteps this difficulty, because in seeing someone's behaviour, we are witnessing their mind in action. Animals, too, may be granted minds, since animals exhibit a range of behaviours from the simple to the complex. In this way, the gulf that Cartesianism would place between us and the rest of the animal kingdom disappears.

However, against its strengths, there are a number of serious obstacles with which analytical behaviourism must contend, as we shall see later.

## 4.2 Analytical contrasted with methodological behaviourism

A much fuller characterization of analytical behaviourism remains to be given. However, before we can proceed it will be necessary to distinguish analytical behaviourism, which is a philosophical position, from behaviourism as it occurs in psychology. If this distinction is not drawn, it may lead to certain confusions, especially among those who have come to philosophy from studying psychology.

Behaviourism in psychology is not a theory regarding how talk about the mind is to be properly understood, but a method of doing psychology, a proposal as to how psychological investigations are to be carried out, hence its full characterization as methodological behaviourism. This approach to psychology is reflected in the fact that the discipline is often defined as the science of human and animal *behaviour*, and not as the study of human and animal *minds*. The idea is that anything that has pretensions to call itself a science must study what is public, because only what is public is capable of measurement and quantification and, even more importantly, verifiable by other observers and experimenters. The results of experiments must be reduplicable by other independent observers if they are to stand any chance of being accepted as viable theories in the so-called 'hard' sciences such as physics, and the same strictures are placed on psychological claims and theories.

Behaviourism in the hands of the American psychologists J.B. Watson and B.F. Skinner took the form of claiming that human behaviour was to be understood as a set of responses evoked by external stimuli. 'Internal' processes, whether in the form of physical brain events or mental events, were set to one side and it was claimed that knowledge of the external stimuli and the behaviour they caused was all that was needed to explain why people behaved as they did. Skinner, in particular, believed that the use of intentional terms in psychology committed one to believing in the existence of a homunculus, a little man in the head, and this leads to the familiar objection that appeals to a homunculus explain nothing, since the abilities and behaviour of the homunculus itself cry out for explanation (see chapter 2, section 2.8.3). Instead, then, of having recourse

to a person's beliefs, desires and intentions in order to explain his behaviour, we are to account for it in terms of conditioned responses which have been reinforced by repetition and reward. A pigeon's behaviour can be 'shaped' by the reinforcement stimulus of a reward of food every time it makes a particular sort of movement, so that ultimately, by this means, it can be made to walk in a figure of eight. In the same way, the whole of human conduct is supposed to be explainable purely in terms of the reinforced stimulus-response relations produced by such operant conditioning.

Having completed this brief outline of methodological behaviourism, we must now return for a closer look at the philosophical variety that is our real goal.

### 4.3 Analytical behaviourism

The analysis of all talk about the mind into statements describing actual and possible behaviour is a form of reductionism, but it represents a different variety from the kind we encountered earlier in relation to the mind/brain identity theory. The mind/brain identity theorists never wanted to maintain that talk about mental states was equivalent in meaning to talk about brain states. Mental concepts were irreducibly different in meaning from physical concepts, of which talk about the brain and its states form a sub-class. But at the level of the items or states of affairs that fall under the concepts, the mind/brain identity theory alleges that there are no separate mental facts over and above physical brain state facts, the mental facts being constituted by the physical facts comprising the material operations of the brain. In the case of the identity theory, the situation is reminiscent of the relation between the temperature of a gas and the kinetic energy of the molecules comprising it. Temperature talk and kinetic-energy-of-molecules talk are not synonymous: it took empirical research to discover the identity of the two apparently different phenomena, not merely reflection upon, and analysis of, the meanings of words. Yet for all that, there are no temperature facts apart from, and over and above, kinetic-energy-of-molecules facts.

The analytical behaviourist, by contrast, insists that statements describing mental or psychological states can be translated, without loss of meaning, into statements describing possible and actual behaviour. The idea is that everything that can be said in the familiar vocabulary of sensations, beliefs, thoughts, consciousness and so forth can somehow be better, or more perspicuously said, in other, non-

mental terms, namely behavioural descriptions. These behavioural descriptions are to be thought of as revealing the true nature of the mental. Analytical behaviourism resembles other forms of reductionism, for example, phenomenalism, which maintains that all talk about physical objects can be rendered in terms of the actual and possible experiences an ideal observer of the objects in question would enjoy. Talk about tables, to take a well-worn illustration of the thesis, is to be translated into statements describing actual and possible table experiences that a suitably placed ideal observer would have. Likewise, reductionism regarding statements about dreams denies that dreams are private experiences which occur to people when they are asleep at night, but are better understood as statements describing their waking experiences in the morning.<sup>2</sup>

It is important to realize, right at the start, that if these patterns of analysis are to stand any chance of success, then the terms comprising the analysis must not contain, or presuppose, any of the mental vocabulary that is being analysed, otherwise the analysis will be circular. In effect, it will merely be repeating itself.

In case this rather abstract point is difficult to grasp at first, I offer a comparison in Exercise 4.1, which you should consider before continuing.

I want to provide an analysis of the concept of a cause, to explain what it means to talk of cause and effect, and, accordingly, I offer the following formula: 'A is the cause of B' is to be analysed as 'B happens because of A'. Can you see why this attempted analysis will not do?

#### Exercise 4.1

I think it is very clear why the analysis given in Exercise 4.1 has to be rejected, because the second statement comprising the alleged analysis is merely a paraphrase of the first. That is to say, it makes use of the very concept of a cause that it is supposed to be analysing. To say 'B happens because of A' is merely another way of saying 'A causes B'. Matters would not be improved if, instead of the original suggestion, I offered 'B is the effect or result of A' Or even 'A makes B occur'. All these, I suggest, carry us no further forward than the statement 'A causes B', with which we began.

A famous analysis of causality by David Hume (1711–76) takes into account the need for the analysis of a concept or a phenomenon to be rendered in genuinely other terms if the analysis is to be informative and to produce a better and clearer understanding. Hume's first shot at 'A causes B' was 'Whenever A occurs, B follows'. Causality, to use Hume's own expression, is the constant conjunction of certain types of event, a regularity that we have discovered on the

basis of repeated experience of one sort of event, in a given set of circumstances, invariably being followed by another sort. We cannot logically deduce an effect from its cause, or a cause from its effect. It is merely that the two sorts of event constituting the causal relation have been found, as a matter of contingent fact, to be related in this way. This has the further important consequence that if two events are causally related, it must be possible to identify each event in complete independence of the other. In other words, there must be descriptions of the events that carry no logical implications for the existence or nature of any other event. We shall return to this important point later in chapter 6, section 6.5, as well as in chapter 7, sections 7.3 and 7.4.

Although this first analysis of causation is not adequate, since it is possible to think of many examples where one sort of event is regularly followed by another sort but one does not cause the other, it is at least genuine and informative and does not make the cardinal error of merely repeating itself.<sup>3</sup>

A shorthand way of conveniently labelling this error, using some Latin tags, is as follows. The phenomenon to be analysed is called the *analysandum*, and the analysis of the phenomenon is called the *analysans*. But, on pain of circularity (and ultimate triviality), the *analysans* must not contain the *analysandum*. (A parallel error can occur with definitions: the *definiens* (the definition of a term) must not contain the *definiendum* (the term to be defined).) The programme of phenomenalism referred to above was ultimately judged by philosophers to be incapable of achievement, because it was discovered that, sooner or later, the *analysans* had to incorporate the very notion of a physical object, the *analysandum*, that it was supposed to be rendering in purely experiential terms.

However, to return to the main discussion, the challenge facing the would-be analytical behaviourist is to produce an analysis of mental terms using a non-mentalistic vocabulary which employs behavioural descriptions alone, containing no expressions that contain, or presuppose, any psychological concepts. But what exactly is a behavioural description?

#### 4.4 Hempel's 'hard' behaviourism

An answer to this question, as well as an understanding of a possible motivation to adopt behaviourism, is supplied by the contribution of Carl Hempel to the debate. Hempel (1905–97), was a member of the

Vienna Circle, the logical positivists, who were chiefly active in the 1920s and 1930s. The logical positivists were much concerned with the nature of scientific knowledge and methodology, and had the achievement of some kind of scientific unity as one of their principal goals. For them, all genuine knowledge and explanation was ultimately to be understood on the model of the physical sciences. Psychology was to be unified with the ‘hard’ science of physics by reduction to it. This conception of knowledge was informed by, and a consequence of, the adoption of the verification principle. According to this principle, unless a statement could be verified empirically (the non-empirical analytic truths of logic and mathematics having been set to one side), it would have to be rejected as devoid of meaning, as literally empty of any significance. The statement in question did not have to be verified in practice, but only in principle. For example, we cannot as yet find out whether it is true that there are planets in the Great Andromeda Nebula because it is too distant. But we know the kinds of thing we would have to do to verify that there are planets there – build a more powerful telescope, send up a space-probe and so forth.

The verificationist theory of meaning had the following consequence. If statements about mental states are construed as concerning logically private states of affairs accessible in principle only to the possessor of those states, as dualism maintains, then statements about the states of minds of others cannot, even in principle, be verified by third persons. But then talk about other minds becomes meaningless. In fact, although one might claim that at least one could verify the existence and nature of one’s own mental states, the positivists would probably disqualify this as genuine verification on the grounds that it is not open to a public, intersubjective check. Given, however, that talk about others’ mental states as well as one’s own, manifestly *is* meaningful, the only way this is achievable is if states of mind are understood to consist in what is indisputably verifiable, namely outward, public behaviour.

It is important, however, to understand what is meant by ‘behaviour’. Look now at Exercise 4.2 before continuing.

Consider, firstly the statements (a) ‘Martin raised his arm’ and (b) ‘Martin’s arm went up’. These are not equivalent in meaning. Can you see the difference in meaning between these two statements?

#### Exercise 4.2

The reason for the non-equivalence of statements (a) and (b) in Exercise 4.2 is as follows. Statement (a) reports something that Martin did, an action he performed, presumably intentionally, and standardly with some purpose in mind. Statement (b) does not report something

Martin did; it merely describes the motion of Martin's arm through space, at best leaving it open whether or not the arm rising was a raising of the arm by Martin, and at worst implying that the motion of the arm was *merely* a motion, i.e. not something that Martin did. Statement (a) entails statement (b), but (b) does not entail (a). If Martin raised his arm, then it follows his arm rose. But if Martin's arm rose, it does not follow that he raised it. His arm might have moved for reasons quite unconnected with his agency. For example, an external reason for his arm moving might be that someone else raised it. Equally, there might have been an internal reason, a muscle spasm perhaps, or the random firing of the nerves that control movement.

This means that there are two distinct modes of behavioural description possible. The first is constituted by descriptions of what people do, of the actions they perform, and the deeds they accomplish through their personal agency. Let us call these, for ease of reference, agential descriptions. The second mode of description characterizes what occurs in terms of bodily movements, often, by implication and context, excluding these from the realm of action, but, at other times, leaving it an open question whether the movements really do comprise actions, or whether they should be viewed as mere bodily movements – 'colourless' bodily movements, to employ the useful expression of the psychologist C.L. Hull. 'Colourless' bodily movement descriptions could, in theory, be refined into sophisticated descriptions of matter in motion, employing the terms and concepts of mathematics, geometry and physics.

Thus, instead of saying, for example, that Suzanne clutched her cheek, we could say instead that a certain piece of matter of such-and-such dimensions (her hand and arm) were observed to move from one set of co-ordinates in space to another set of co-ordinates over a specified period of time. Similarly, rather than saying that someone smiled, we would have to say instead that a piece of flesh of certain dimensions underwent changes in its shape, characterizing these alterations in the language of mathematics and topology.<sup>4</sup>

You should now look at Exercise 4.3 before reading on.

#### Exercise 4.3

Can you see what mode of description Hempel must choose for his behavioural analysis and why? Give reasons for your decision.

At the risk of importing circularity into his proposed analysis, Hempel must confine himself to using descriptions of colourless bodily movements. This would also accord with his avowed aim of reducing psychology to physics. In line with this, Hempel wrote: 'We

see clearly that the meaning of a psychological statement consists solely in the function of abbreviating the description of certain modes of physical response characteristic of the bodies of men and animals.<sup>5</sup> It has to be said, however, that Hempel's attempted behaviourist analysis, when it appeared, of a psychological statement, 'Paul has a toothache', was less of a success, instantly violating the stricture that no analysis should contain the very terms it is supposed to be analysing. Hempel's proposed analysis comprised five elements:<sup>6</sup>

- 1 Paul weeps and makes gestures of such and such kinds.
- 2 At the question 'What is the matter?', Paul utters the words 'I have a toothache'.
- 3 Closer examination reveals a decayed tooth with exposed pulp.
- 4 Paul's blood pressure, digestive processes, the speed of his reactions, show such and such changes.
- 5 Such and such processes occur in Paul's central nervous system.

Conditions (3), (4), and (5) do not appear to be essential to an analysis of the meaning of the statement 'Paul has a toothache'. It could be true that Paul had a toothache, even if a physical examination revealed no decayed tooth and no changes in his blood pressure, digestive processes or nervous system. Conversely, the physical changes just described might occur, and yet Paul feels no pain. To be sure, this might be puzzling, but it would not be self-contradictory to conjoin Paul's truthful report that he had toothache with the denial that anything out of the ordinary was happening to his teeth or body. (After all, think of people who have so-called psychosomatic pains for which no obvious physical cause is forthcoming.) This is sufficient to show that (3), (4) and (5), cannot be part of what it means for Paul to have toothache.

That leaves (1) and (2). In relation to (1) the problem is that terms such as 'weeps' and 'gestures' already imply the attribution of mental states to Paul. 'Weeps' implies that Paul is suffering some unpleasant experience (a more neutral description would be 'water is coming out of Paul's eyes'), and 'gestures' suggests hand movements made by Paul with the intention of indicating the place and intensity of the pain, and perhaps to get help and sympathy.

In relation to (2), a further difficulty that infects the analysis is that it is clear that Paul cannot respond to the question unless he understands it. Moreover, he must affirm the words that come out his mouth, that is to say, he must know what the words mean and intend them to answer the speaker's question.<sup>7</sup> But understanding, knowing,

affirming and intending are all behaviourally unanalysed mental terms. Lastly, Paul will only respond to the question in the way he does, if he wants to tell the truth. It might be that Paul is suffering, but not from toothache, and he wishes to conceal the real reason for his distress. But ‘wanting’ is again a mental or psychological term which the behavioural analysis is supposed to be eschewing in order to avoid the charge of circularity.

## 4.5 Specifying patterns of behaviour

A serious attempt to carry out the analysis of mental terms in a non-circular way has to face a number of difficulties. The first problem concerns which patterns of behaviour, characterized purely as bodily movements, the motion of matter through space, are to comprise the analysis. The root of the problem lies in the fact that there is no neat one-to-one correspondence between types of action and types of bodily movement. This is reminiscent of the difficulty that confronted the type-type mind/brain identity theorists: a given type of mental state might be identical with C-fibres firing in Keith’s brain, but Z-fibres firing in Zielfa’s brain or even P-fibres firing in Pippa’s brain. Although every token of a type of mental state is identical with a token of some type of brain state, it is not necessarily one and the same type of brain state on every occasion: a token of a given type of mental state is multiply realizable in different types of brain state. So it is with actions. Every token of a type of action will be identical with a token of some type of bodily movement (movement includes stillnesses), but it is not necessary that each token of a given type of action must find its embodiment in exactly the same type of bodily movement on every occasion when it occurs.

Before reading on, consider the question in Exercise 4.4.

### Exercise 4.4

Can you see the threat posed by the multiple realizability of actions to the behavioural analysis of mental states?

The threat suggested in Exercise 4.4 is that the analysis could never be carried through to completion. To see this, consider again the first two items in Hempel’s purported behavioural analysis of ‘Paul has a toothache’. Hempel translates this into ‘Paul weeps, makes various sorts of gesture, and when asked what is wrong, utters the words “I have a toothache”’. Well, Paul *may* do these things, but there is an indefinite variety of other things he might do instead which could

equally count as behavioural expressions of the pain he feels. He might, for example, shout out, scream, reach for the whisky bottle, apply oil of cloves, sit there in grim silence or thump the bed. If he is a monoglot Frenchman, he will not utter the English words 'I have a toothache', but the French equivalent. On the other hand, if he is a Filipino, he might simply moan 'Agoy!' Remember: if the analysis is to be successful, it must mention all those types of behaviour, and only those types of behaviour, that are capable of constituting Paul's expression of pain. This suggests that, following Hempel's proposal, the analysis will look something like this:

Paul has the toothache = *Either* Paul is groaning *or* he is wincing  
*or* tears are coming out of his eyes *or* he  
 is uttering the words 'I am in pain' *or*  
 their various foreign equivalents *or* he  
 is reaching for the whisky bottle *or* . . .

There are at least two things wrong with this analysis. Firstly, it appears that it could never be completed: there is an indefinite variety of things Paul might do. Secondly, the items on this list mention actions. But descriptions of actions are terms that implicitly refer to mental states and are prohibited from inclusion in the analysis. This means that each type of action-description in turn will itself need to be cast into a lengthy and complex disjunction of the various types of bodily movement through which the action could theoretically manifest itself. Thus it strongly appears that the analysis could never be carried through to a final conclusion.

Consider also that if the problem of specifying which behaviours are to constitute the analysis of a simple report of a sensation such as toothache looks hard, consider how much more difficult such an analysis would be to carry out when it needs to refer to the behavioural manifestations of a complex and abstract belief. Suppose I believe that no one as yet has been able to explain what consciousness is, and why it exists. How should this belief be analysed in behavioural terms? One obvious way, perhaps, is through my verbal expression of the belief by means of the utterance: 'I believe that no one as yet has been able to explain what consciousness is, and why it exists.' But we have already noted that such an analysis covertly contains mental items and therefore has to be rejected. Most importantly, however, it should be clear that there is a massively indefinite variety of ways in which I, or someone else, might express the belief about the intractability of providing an account of consciousness, and this means that the prospect of achieving a successful analysis recedes even further.

However, according to John Foster, this is not necessarily a conclusive objection. It may only show that 'in the case of at least some (and perhaps all) psychological statements, the linguistic items which make explicit their ultimately non-mentalist content are of infinite length, thus showing that the content itself, in its fully analysed form, is of infinite complexity'.<sup>8</sup>

In so far as I understand him on this matter, what Foster is maintaining is that this does not mean that the notion of a non-mentalist analysis of mental items is incoherent, just because, by its very nature, it could never be encompassed by us. All it shows ultimately, perhaps, are that there are things that we can say by means of a mentalistic vocabulary that we cannot manage to say purely by the use of a putative physicalist translation of that vocabulary.

Another obstacle to the analysis we need to consider is this: how are we to identify which types of colourless bodily movements are to be included in it, and which are to be excluded? There is nothing in the description of colourless bodily movements as such which provides a clue to their possible inclusion or exclusion in any particular pattern of behavioural analysis of a mental item. In other words, considered in themselves from the point of view of physical science, the various kinds of bodily movement form no natural or obvious grouping, such that we can read off from them which are to be included and which excluded from a behavioural translation of a particular mental state. It would appear that their identification can only be approached through the identification of the possible set of actions selected to comprise the initial stage of the analysis. The idea is this: think of which set of actions most plausibly could serve as the behavioural manifestations of a particular type of mental state. Then, reflecting upon these actions, try to imagine the gross bodily movements that would need to figure in their execution and those that would be excluded. For example, consider the action 'scoring a goal by kicking the ball into the net'. Movements of the leg of various types could figure in the analysis but not, say, movements of the head. However, what this is leading up to is that the identification of the movements has to proceed via the identification of the action. It is the action that gives the group of movements their unity, and it is only via the action that the grouping is identifiable and intelligible. But this means that the analysis relies upon an indispensable reference to a mentalistic item for its execution, and it was precisely all reference to such items that was to be excluded from the analysis on pain of circularity.<sup>9</sup>

## 4.6 Circularity and infinite regression

The actions we perform do not result from single mental states, but combinations of mental states. For example, suppose Manjit believes that the shop across the road contains cigarettes, but he has just given up smoking. Clearly, he will not cross the road with the intention of entering the shop and purchasing the cigarettes. On the other hand, if he hasn't given up smoking and wants to smoke, then this, in combination with his belief that cigarettes are to be found in the shop, will lead to him crossing the road.

With this in mind, let us suppose we want to analyse the mentalistic statement, 'Manjit wants a packet of cigarettes', behaviourally. How will the analysis proceed? We might try rendering this by saying that Manjit will go and buy some cigarettes. But clearly this is too simple. He won't do this if his intention to give up smoking is stronger than his desire to smoke, or if he believes he hasn't sufficient money to pay for the cigarettes, or that the shop is shut or temporarily out of cigarettes and so on. So it would appear that any attempt to spell out Manjit's desire for cigarettes purely in behavioural terms will involve a reference to other mental states, in this case a number of beliefs. But, fairly obviously, the analysis cannot be allowed to contain unanalysed mental terms such as 'belief', and some way of fleshing out what beliefs amount to in terms of behaviour will have to be found. The problem is that it appears that any attempt to spell out how Manjit is likely to behave, given that he believes the shop sells cigarettes, must involve a reference to Manjit's desires. His belief alone cannot motivate him to action, i.e. to behaving in a certain way. If Manjit is to cross the road and buy cigarettes, the belief needs to be combined not only with other beliefs but also with a desire for cigarettes. In fact, the matter is more complicated still, because Manjit's desire to buy cigarettes will only lead to action providing it is stronger than his other desires, such as the desire not to get knocked down crossing a busy street, or the desire to avoid crossing over into hot sunlight. Thus it appears that a behavioural analysis of a desire must involve a reference not only to other desires, but particularly to beliefs viewed as unanalysed mental items. And, likewise, a behavioural analysis of a belief must involve a reference not only to other beliefs but also to desires regarded as unanalysed mental items.

There are two circularities involved here, a larger one and a smaller one. The larger circularity is that no behavioural analysis

may be allowed to contain unanalysed mental terms, and the problem is that however far we go in attempting to cash out mental items in terms of behavioural descriptions, a residue of unanalysed mental items will always be left, crying out for yet more behavioural analysis. It appears that this process will be unending, though arguably this does not ultimately compromise the analysis if Foster is correct.<sup>10</sup>

The smaller circularity consists in the fact that a behavioural analysis of a desire has to make use of the unanalysed notion of a belief, and the behavioural analysis of a belief has to make use of the unanalysed notion of a desire. Manjit's desire for a cigarette consists in, among other things, his crossing the road to the shop, given that he believes it sells cigarettes; and his belief that the shop sells cigarettes consists in, among other things, his crossing the road to the shop given that he desires a cigarette. In other words, the explanation of what it is for Manjit to want a cigarette involves a reference to Manjit's beliefs, and the explanation of what it is for Manjit to hold a belief about the availability of cigarettes involves a reference to Manjit's desires. Each half of the analysis takes in, as it were, the other half's washing.

However, it is arguable that this smaller circularity can be avoided. Sydney Shoemaker argues that the problem parallels the case of someone trying to give an account of what positive and electric charges are:

Imagine someone whose beliefs about electric charge are summed up in the following sentence: 'Things with positive charge attract things with negative charge and repel things with positive charge; things with negative charge attract things with positive charge and repel things with negative charge; and negative charge can be induced in a rubber comb by rubbing it against wool.' And suppose he is asked to define 'positive charge' and 'negative charge' in terms of this set of beliefs. It might seem that there is no way he could do this without running into circularity, since each sort of charge is characterized in terms of its relation to the other.<sup>11</sup>

The solution, says Shoemaker, following a proposal by David Lewis, is to make use of the notion of a 'Ramsey sentence', so-called because it was invented by the Cambridge philosopher Frank Ramsey (1903–30). (We shall encounter the use of a Ramsey sentence again when we consider functionalism in chapter 5, section 5.5, where it is put to a similar purpose.)

A Ramsey sentence is constructed like this. The terms 'positive charge' and 'negative charge' are eliminated by being substituted

by the terms ‘properties F and G’. The original attempt to define positive and negative charge thus transforms into: ‘For some properties F and G, things with F attract those with G and repel things with F; things with G attract things with F and repel things with G; and G can be induced in a rubber comb by rubbing it with wool.’

The phrase, ‘For some properties F and G’, which prefaces this definition, is another way of expressing what logicians call the existential quantifiers  $\exists(F)$  and  $\exists(G)$ , which are to be read respectively as ‘There are some things F such that . . .’, and ‘There are some things G such that . . .’. F and G do not mention the properties by name, and hence circularity can be avoided, since the terms being defined do not occur in the *definiens* of each definition (see section 4.3 if you have forgotten the meaning of this term).

## 4.7 Ryle’s ‘soft’ behaviourism

In this section I want to say something about the importance of dispositions in analytical behaviourism, and I will approach this topic through a brief exploration of the work of Gilbert Ryle (1900–76), with whom the notion of a disposition is particularly associated. Ryle’s behaviourism, in so far as it can properly be classified as behaviourism, is to be found in his book *The Concept of Mind*, first published in 1949.<sup>12</sup>

The motivation of Ryle’s account of the nature of the mind is very different from Hempel’s. Firstly, there is no evidence in Ryle’s book that he subscribed to the verificationist theory of meaning outlined earlier. This is fortunate, as subsequent philosophical work on this issue has shown that verificationism is a dubious account of semantic content, fatally confusing being able to specify which state of affairs, if it occurred, would make a factual statement true with the quite different requirement of being able to tell whether or not the state of affairs in question does so obtain. Only the first of these two requirements can reasonably be viewed as spelling out what is essential for a factual statement to have significance.

Secondly, Ryle was no physicalist seeking to reduce psychology to physics. He was as suspicious of physicalist theories of the mind as he was contemptuous of Cartesian dualism, seeking to dissolve, rather than solve, the mind/body problem, by approaching the issue in a new way: ‘[T]he hallowed contrast between mind and matter will be dissipated, but dissipated not by either of the equally hallowed

absorptions of Mind by Matter or of Matter by Mind, but in a quite different way.<sup>13</sup>

In the introduction to his book, Ryle states that his avowed aim is to determine the logical cross-bearings of concepts of the mental, to enable people who can already talk sense *with* these concepts to be able, in addition, to talk sense *about* them, after the manner of the logical or philosophical map-maker who seeks to gain a synoptic view of the concepts by making clear their interrelations and the regulations governing their uses.

What this means, in practice, is that it is Ryle's aim to demolish the Cartesian conception of the mind as a ghostly non-physical entity existing over and above the familiar flesh-and-blood living human being, an entity whose states are supposed to be logically private in the sense explained earlier (chapter 1, section 1.5) to the individual whose mind is in question. However, it would be perfectly possible to repudiate dualism, without thereby being forced to abandon belief in the radical privacy of mental states, even while accepting that a commitment to dualism entails a commitment to acceptance of the privacy of the mental. Consequently, I shall consider Ryle's attack on these twin aspects of dualism separately.

Ryle's first aim is to repudiate utterly the Cartesian concept of mind as an immaterial substance linked in life to a corporeal machine, the body. Ryle characterizes this with, as he says himself, 'deliberate abusiveness', as 'The Dogma of the Ghost in the Machine.' It is, Ryle claims,

entirely false, and false not in detail but in principle. It is not merely an assemblage of particular mistakes. It is one big mistake and a mistake of a special kind. It is namely a category mistake. It represents the facts of mental life as if they belonged to one logical type or category (or range of types or categories), when they actually belong to another.<sup>14</sup>

The mistake Ryle is thinking of is this. There are not, in addition to living human beings whose outward behaviour and inner workings are as public and open to scrutiny as the careers of trees, crystals and planets, ghostly entities called minds, whose inner 'paramechanical' workings are accessible only to the persons to whom those minds happen to belong. A person does not live through two collateral histories, the one consisting in outward public physical doings, and the other consisting in ghostly happenings on a private mental stage.

To illustrate the point he is making, Ryle offers what has become a famous analogy. We are to imagine a foreign visitor being shown

around Oxford University. He visits the various colleges, the Bodleian Library, the laboratories, the Examination Schools in the High Street, the sub-faculty of Philosophy library and so forth. At the end of his tour, he says, 'Thank you very much. But unfortunately I have not yet been shown the university itself.' His mistake is obvious. He is thinking of the university as if it were a building itself, like the individual colleges or the examination school, and capable of existing independently of them, rather than realizing that the totality of what he has seen *is* the university.

The same mistake is evident, according to Ryle, in supposing that, in addition to the overt public performances of human beings, the purposeful, intelligent actions they carry out, and operations they engage in, there is something else, a non-physical entity called the mind, existing over and above the body, which is the true, hidden repository of a person's intellectual and mental repertoire.

In accordance with this, Ryle writes, in what is a pretty typical passage: 'Overt intelligent performances are not clues to the workings of minds; they *are* those workings. Boswell described Johnson's mind when he described how he wrote, talked, ate, fidgeted and fumed.'<sup>15</sup> In other words, for Ryle, Johnson's behaviour is not evidence that something is going on behind it in the secret recesses of Johnson's mind. Rather, Johnson's mental processes are manifest in his behaviour. His overt intelligent performances are not a substitute for workings of his mind, a mere stand-in, or proxy. Instead, they *are* those workings, fully open to public view and inspection.

It is thus not difficult to appreciate why Ryle should have been labelled a behaviourist. If mental processes are not ghostly concealed processes in an immaterial soul, and if the workings of a person's mind are identified with overt public performances, the reduction of mind to matter also having been repudiated, it is difficult to see what else the facts about a person's mental life could amount to, if not facts about his behaviour. (Another option would be functionalism – see chapter 5 – but this option was not open to Ryle, as the theory had not yet been invented at the time he was writing.)

Notice, however, that Ryle's behaviourism is a far cry from Hempel's ideal of rendering all descriptions of behaviour in terms of colourless bodily movements and the concepts of physics. Ryle is quite clearly not in the least worried that his analysis will be accused of circularity, because he is not attempting to reduce mental states to purely physical descriptions after the manner of the logical positivists. This is why Ryle's variety of behaviourism, by contrast with the 'hard' behaviourism of Hempel, has been labelled 'soft' behaviourism.<sup>16</sup>

## 4.8 Ryle and dispositions

Nothing very much has been said so far about the need to include descriptions not only of actual, but of potential behaviour in any proposed behavioural analysis of mental states, and to this we now turn. It would be highly implausible to maintain the thesis that mental states consist entirely in actual overt public behaviour, because this fails to accommodate the commonplace fact that people can be in all kinds of mental state without ever revealing these through their behaviour. For instance, at this moment there are all kinds of beliefs and intentions that I harbour, but someone observing me would not be able to tell exactly what beliefs and intentions these are. Even if I had a toothache, this would not necessarily reveal itself in my behaviour. For example, if I am at a concert and anxious not to spoil the enjoyment of others, I might well sit there and suffer in silence, no one being able to ascertain from my composure the agonies I am enduring.

To avoid this difficulty, the notion of a disposition to behave was invoked. This means that the behavioural analysis has to include not merely actual behaviour, but potential behaviour as well. For me to be described truly as having the toothache, it is not necessary that I should actually be groaning, clutching my cheek and so forth. Rather, it is enough that I am disposed to groan, clutch my cheek, etc. in the appropriate circumstances. In other words, it is to say that if certain conditions were fulfilled, then I would groan. The same applies to intentional states such as beliefs and desires. To describe me as believing that the moon is made of green cheese is not, improbably, to commit oneself to saying I actually am avowing this belief, but only that, in the right circumstances, I would avow it.

Ryle's way of making the point is to say that to have a belief is to have a tendency or to be prone to act in a certain way, and he goes on to explain this by using the examples of brittleness and solidity. To say that glass, for example, is brittle, is not to say it actually is shattering, but only that if it were struck, then it would shatter. Similarly, to describe sugar as soluble is to say that if it were placed in water, then it would dissolve. According to Ryle, for something to have a disposition is simply for a whole lot of hypothetical 'if . . . then . . .' statements to be true. In line with his denial of mental states as events different in kind from, and concealed behind, behaviour, Ryle is anxious to emphasize that to have a disposition is *not* to be in a certain kind of state: 'To possess a dispositional property is not to be in a particular state, or to undergo a particular change; it is to be

bound or to be liable to be in a particular state, or to undergo a particular change when a particular condition is realized.’<sup>17</sup> Before reading on, look at Exercise 4.5.

How plausible do you find Ryle’s claim that having a disposition consists simply in a set of hypothetical ‘if . . . then . . .’ statements being true, providing the appropriate conditions are realized? Consider your response and discuss it with someone else before reading my comments.

#### Exercise 4.5

Most philosophers would now agree that there is more to a disposition than what Ryle claims it to be. The reason that glass has a disposition to break when struck – the explanation of its brittleness – is to be found in facts about its underlying micro-structure. If the micro-structure were to be altered in an appropriate manner, as when glass is annealed, then its tendency to shatter when struck is much reduced, or even removed, as in the case of toughened glass. By parity of argument, to have a belief (or any other mental state) is not merely to be inclined to behave in a certain way, but to be in an inner state, which may, in the appropriate circumstances, manifest itself in outward behaviour. However, once the propriety of thinking of beliefs and other mental states as inner causes of outward behaviour has been allowed, behaviourism has been abandoned.

This critique of Ryle’s refusal to countenance dispositional mental states as consisting in anything more than tendencies to behave in particular ways, leads to another important criticism of behaviourism. The essence of the objection is this: it is a commonplace that we explain why people exhibit the behaviour they do in terms of their states of mind. For example, a sharp pain makes me cry out. The realization that I’ve left my keys in the door leads me to turn round and go home. The desire for some cigarettes, coupled with the belief that the shop over the road sells them, explains why I cross the road.

But if the pain, the realization and the desire are all to be rendered in purely behavioural terms, then how can they be appealed to in order to explain my behaviour, since this will amount to attempting to explain one bit of behaviour in terms of another bit. It will be tantamount to the futile task of trying to explain behaviour by reference to itself. Items that explain behaviour cannot, so to speak, be on the same level as the behaviour they are invoked to explain. Rather, they must be conceived of as the underlying causes of the surface behaviour in question, i.e. they must themselves be non-behavioural in nature.

The point was well made by Hilary Putnam when writing about what we mean when we attribute a disease such as multiple sclerosis

to someone. It is plausible to say that 'Normally people who have multiple sclerosis have some or all of the following symptoms' is a necessary (analytic) truth. In other words, it is part of what we mean by 'multiple sclerosis' that normally it gives rise to certain symptoms. But, it must be emphasized, this does not mean that talk about the disease, multiple sclerosis, can be translated, without loss of meaning, into talk about the symptoms of multiple sclerosis. This would be analogous to the error the analytical behaviourists make when they try to render talk about 'pain' in terms actual and potential pain-behaviour. The error that underpins both these attempts is the mistake of trying to make *causes* logical constructions out of their *effects*. But pains and diseases are causes of behaviour and symptoms respectively, and cannot be reduced to, and constructed out of, their effects.<sup>18</sup>

#### **4.9 The denial of the subjective 'inner' features of mental states**

One of the strongest objections to behaviourism derives from outraged common sense. Because of its insistence on reducing mental states to patterns of outward behaviour, or dispositions to engage in such behaviour, behaviourism effectively denies the existence of the 'inner' aspects of mental states, their qualia. But obviously there is more to having toothache than groaning and clutching one's cheek, or even being disposed to do so. When one goes to a dentist or takes some aspirin, the aim is not merely to prevent oneself from behaving, or being disposed to behave, in a certain way, but to get rid of the unpleasant sensation, the pain one is experiencing. It is for this reason that behaviourists were accused of feigning anaesthesia. Accounting for sensations in behavioural terms comprises a particularly intractable difficulty, because it is of the essence of sensations to possess a phenomenology. In short, there is something it feels like to have an itch, or pins and needles in your foot, or an ache in the shoulder. These all involve consciousness: there is something it is like to undergo such experiences.

The experience of visualizing something in 'the mind's eye' is also peculiarly difficult to account for in behaviourist terms. We unhesitatingly think of this experience as consisting in private mental imagery, which frequently does not, and does not need to, manifest itself in outward behaviour. But in his anxiety to repudiate the existence of inner states, Ryle writes:

[A] person picturing his nursery is, in a certain way, like that person seeing his nursery, but the similarity does not consist in his really looking at a real likeness of his nursery, but in his seeming to see his nursery itself, when he is not really seeing it. He is not being a spectator of a resemblance of his nursery, but he is resembling a spectator of his nursery.<sup>19</sup>

Before reading my response to Ryle's account, look at Exercise 4.6.

How plausible do you find Ryle's account of what it is to picture one's nursery to oneself? What criticisms can you make of this account?

Exercise 4.6

What Ryle is maintaining is thin and unconvincing. To begin with, how could I resemble a spectator of my nursery if I were lying in bed, undressed, in the dark, engaging in a spot of nostalgia, by visualizing my nursery to myself? Besides, as I have already remarked, the experience of visualizing my nursery is a prime example of the sort of case where we would want to say that a private episode of consciousness, not capturable in terms of outward public behaviour, is integrally involved.

## 4.10 The denial of first-person knowledge and authority regarding mental states

We allowed in the first chapter (section 1.7) that there are occasions when it seems that others know our own minds better than we do ourselves, temporarily requiring us to adopt a third-person perspective from which to make an assessment of our state of mind. But for most of the time, and especially as regards non-intentional states such as sensations, we are in the best position to say what we are feeling or thinking. Moreover, we can describe our sensations, or avow our beliefs or intentions, without having to go through any process of finding out what these are through observation or inference from more basic data. In particular, I do not need to observe my own behaviour in a mirror or on a video in order to be able to say what I believe or how I feel. Behaviourism, by making my mental states a matter of how I behave, reverses all this, and leads to jokes such as: 'One behaviourist meets another on the street. You feel fine, he says to the other. How do I feel?' Or one behaviourist says, after making love to another: 'It was great for you. How was it for me?' To this, we

may add the further observation that if there were nothing more to being in pain or holding a belief than behaving or being disposed to behave in a certain way, we would have the absurd consequence that a person would have to wait until he or she exhibited the appropriate behaviour before being able to report his or her state of mind.

#### **4.11 Can the possibility of pretence show that behaviourism cannot be correct?**

It might be said that behaviourism cannot be correct because a person could be merely pretending to feel pain. That is to say, there might be plenty of pain-behaviour but no corresponding pain. How then can pain be identical with pain-behaviour, since one can occur without the other? The behaviourist, however, has a comeback. The point of the example is to show that there can be behaviour without a corresponding mental state – in this case pain – and hence there must be more to mental states than merely behaviour. However, whilst there are independently good reasons for thinking this conclusion to be true, it does not seem that it can be established on the basis of the possibility of pretence. Pretence implies the presence of an intention to pretend, and this intention, the behaviourist might argue, can be analysed into the mock pain-behaviour that constitutes the pretence. To put this another way, it is true that the pain-behaviour cannot be identified with pain, for the person is feeling none. But the pain-behaviour can still be identified with *some* state of mind, the behaviourist can contend, namely the more complex mental state of intending to pretend. Hence the argument from the possibility of pretence cannot show what it was intended to show, namely that, despite an abundance of behaviour, no mental states are present, and thus behaviour cannot constitute mental states, contrary to what behaviourism maintains.

The possibility of zombies, however, which was discussed briefly in chapter 3, section 3.8.8, is sufficient to establish the falsity of behaviourism. Zombies are physically just like ordinary human beings, except that they lack a mental life. They behave to all intents and purposes just like human beings, but they lack any consciousness or knowledge of the motions their bodies are making. In other words, zombies exhibit all the right sorts of behaviour as far as an external observer can tell, but, nevertheless, mentality is entirely absent. This does appear to be an intelligible fantasy, unvitiated by internal

contradiction, and if this is indeed so, then behaviourism has to be rejected on this ground alone.<sup>20</sup>

The falsity of behaviourism would also be demonstrated by the converse possibility, namely the presence of mentality in the absence of any behavioural manifestations. Hilary Putnam has devised the following thought-experiment to provide a convincing case for this possibility. We are to imagine a race of super-Spartans or super-Stoics who constitute a community in which the adults have the ability to suppress successfully all involuntary pain-behaviour. Although they sometimes admit that they are in pain, they always do this in calm, well-modulated voices, even if they are suffering the torments of the damned. Neither do they groan, scream, wince, sweat nor grit their teeth. They admit it takes a great effort of will to do this, but they have important ideological reasons for behaving in the way they do, and they undergo years of training to achieve the right standard of behaviour in the face of pain. It might be argued that the lack of pain-behaviour will only be found in adults who have been suitably conditioned, and that non-adults will exhibit unconditioned pain-reactions, thus giving the ascription of mental states to the community a toehold. However, in response to this objection Putnam elaborates the fantasy further. We are to suppose that after several millions of years, the super-Spartans begin having children who are born fully acculturated, speaking the adult language, sharing opinions about their society and also the beliefs about not giving any signs that one is in pain, unless it is by verbal report in a calm, unanxious and unconcerned manner. There are, then, no unconditioned pain-responses in this community, yet it would be crazy to take the position that because there is a total absence of natural, unconditioned pain-behaviour, the super-Spartans cannot have the capacity for experiencing pain to them.

Putnam also asks us to imagine super-super-Spartans, who inhabit the X-world. These people have been super-Spartans for so long that not only do they not evince unconditioned pain-behaviour, but they have begun to suppress all talk about pain. X-worlders do not admit to having pains, and pretend not to know either the word 'pain' or the experience to which it refers. Nevertheless, they do have pains and know that they do. Putnam concludes:

If this last fantasy is not, in some disguised way, self-contradictory, then logical behaviourism is simply a mistake. Not only is the second thesis of logical behaviourism – the existence of a near-translation of pain into behaviour talk – false, but so is even the first thesis – the existence of 'analytic entailments'. Pains are responsible for certain kinds of

behaviour – but only in the context of our beliefs, desires, ideological attitudes, and so forth. From the statement ‘X has a pain’ by itself no behavioural statement follows – not even a behavioural statement with a ‘normally’ or a ‘probably’ in it.<sup>21</sup>

The most extreme example of mental states without behaviour is provided by the possibility of disembodied existence. It might be argued that only a limited range of experiences could be ascribed to disembodied persons, since they would not have bodies in which to feel pains, or eyes with which they could see. But there is no reason why it should not seem to such people that they had bodies and eyes, even though they did not, and that it seemed to them that they were feeling pains in these imaginary bodies and were having visual experiences as a result of the operations of their imaginary eyes. It is arguable that Descartes’ argument from clear and distinct perception (see chapter 2, section 2.7.2), as well as Kripke’s argument against the identity theory (see chapter 3, section 3.8.8), do succeed in establishing the possibility of disembodied mental states – that is, that even if such states cannot exist in this world, there is a possible world in which they do, and this is all that is needed to prove that analytical behaviourism is false.

Questions  
to think  
about

- 1 In your opinion, does behaviourism offer a clear and satisfactory solution to the mind/body problem?
- 2 Can behaviourism account for the existence of an inner mental life with no behavioural manifestations?
- 3 Could there be super-Spartans who feel pain but never exhibit pain-behaviour?
- 4 Do zombies represent a genuine possibility?
- 5 When I say ‘I have an itch’, am I describing something that is private to me?

### Suggestions for further reading

- A clear and excellent introduction to analytical behaviourism can be found in Stephen Priest’s book, *Theories of the Mind* (London: Penguin, 1991).
- An even brisker account is available in Paul Churchland’s *Matter and Consciousness* (Cambridge, MA: MIT Press, 1984).
- One of the very best discussions can be found in Jaegwon Kim’s *Philosophy of Mind* (Boulder, CO: Westview Press, 1996). This is very useful for beginning students who wish to go further.
- John Foster’s treatment of behaviourism in his book *The Immaterial Self*:

*A Defence of the Cartesian conception of the Mind* (London: Routledge, 1991) is extremely rigorous and thorough. Not recommended for beginners, but very useful at a later stage.

- Peter Smith and O.R. Jones provide a stimulating and clear discussion of behaviourism in their book *The Philosophy of Mind* (Cambridge: CUP, 1986). It is highly recommended, especially for beginners.