

## The puzzle of representation

When NASA sent the Pioneer 10 space probe to explore the solar system in 1972, they placed on board a metal plate, engraved with various pictures and signs. On one part of the plate was a diagram of a hydrogen atom, while on another was a diagram of the relative sizes of the planets in our solar system, indicating the planet from which Pioneer 10 came. The largest picture on the plate was a line drawing of a naked man and a naked woman, with the man's right hand raised in greeting. The idea behind this was that when Pioneer 10 eventually left the solar system it would pursue an aimless journey through space, perhaps to be discovered in millions of years time by some alien life form. And perhaps these aliens would be intelligent, and would be able to understand the diagrams, recognise the extent of our scientific knowledge, and come to realise that our intentions towards them, whoever they may be, are peaceful.

It seems to me that there is something very humorous about this story. Suppose that Pioneer 10 were to reach some distant star. And suppose that the star had a planet with conditions that could sustain life. And suppose that some of the life forms on this planet were intelligent and had some sort of sense organs with which they could perceive the plate in the spacecraft. This is all pretty unlikely. But even having made these unlikely suppositions, doesn't it seem even *more* unlikely that the aliens would be able to *understand* what the symbols on the plate mean?

Think about some of the things they would have to understand. They would have to understand that the symbols on the plate *were* symbols – that they were intended to stand for things, and were not just random scratches on the plate, or mere decoration. Once the aliens knew that they were symbols, they would have to understand what sort of symbols they were: for example, that the diagram of the hydrogen atom was a scientific diagram and not a picture. Then

## The puzzle of representation

they would have to have some idea of what sorts of things the symbols symbolised: that the drawing of the man and woman symbolised life forms rather than chemical elements, that the diagram of the solar system symbolises our part of the universe rather than the shape of the designers of the spacecraft. And – perhaps most absurd of all – even if they did figure out what the drawings of the man and woman were, they would have to recognise that the raised hand was a sign of peaceful greeting rather than of aggression, impatience or contempt, or simply that it was the normal position of this part of the body.

When you consider all this, doesn't it seem even more unlikely that the imagined aliens would understand the symbols than that the spaceship would arrive at a planet with intelligent life in the first place?

One thing this story illustrates, I think, is something about the philosophical problem or puzzle of representation. The drawings and symbols on the plate represent things – atoms, human beings, the solar system – but the story suggests that there is something puzzling about how they do this. For when we imagine ourselves into the position of the aliens, we realise that we can't tell what these symbols represent just by looking at them. No amount of scrutiny of the marks on the plate can reveal that these marks stand for a man, and these marks stand for a woman, and these other marks stand for a hydrogen atom. The marks on the plate can be understood in many ways, but it seems that nothing in the marks *themselves* tells us how to understand them. Ludwig Wittgenstein, whose philosophy was dominated by questions about representation, expressed it succinctly: 'Each sign *by itself* seems dead; *what* gives it life?'.<sup>8</sup>

The philosophical puzzle about representation can be put simply: how is it possible for one thing to represent something else? Put like this, the question may seem a little obscure, and it may be hard to see exactly *what* is puzzling about it. One reason for this is that representation is such a familiar fact of our lives. Spoken and written words, pictures, symbols, gestures, facial expressions can all be seen as representations, and form the fabric of our everyday life. It is only when we start reflecting on things like the Pioneer 10 story that

## The puzzle of representation

we begin to see how puzzling representation really is. Our words, pictures, expressions and so on represent, stand for, signify or mean things – but how?

On the one hand, representation comes naturally to us. When we talk to each other, or look at a picture, what is represented is often immediate, and not something we have to figure out. But, on the other hand, words and pictures are just physical patterns: vibrations in the air, marks on paper, stone, plastic, film or (as in Pioneer 10) metal plates. Take the example of words. It is a truism that there is nothing about the physical patterns of words themselves which makes them represent what they do. Children sometimes become familiar with this fact when they repeat words to themselves over and over until they seem to ‘lose’ their meaning. Anyone who has learned a foreign language will recognise that, however natural it seems in the case of our own language, words do not have their meaning *in and of themselves*. Or as philosophers put it: they do not have their meaning ‘intrinsically’.

On the one hand, then, representation seems natural, spontaneous and unproblematic. But, on the other hand, representation seems unnatural, contrived and mysterious. As with the concepts of time, truth and existence (for example) the concept of representation presents a puzzle characteristic of philosophy: what seems a natural and obvious aspect of our lives becomes, on reflection, deeply mysterious.

This philosophical problem of representation is one main theme of this book. It is one of the central problems of current philosophy of mind. And many other philosophical issues cluster around this problem: the place of the mind in nature, the relation between thought and language, the nature of our understanding of one another, the problem of consciousness and the possibility of thinking machines. All these issues will be touched on here. The aim of this chapter is to sharpen our understanding of the problem of representation by showing how certain apparently obvious solutions to it only lead to further problems.

## The puzzle of representation

### The idea of representation

I'll start by saying some very general things about the idea of representation. Let's not be afraid to state the obvious: a representation is something that represents something. I don't say that a representation is something that represents something *else*, because a representation can represent itself. (To take a philosophically famous example, the 'Liar Paradox' sentence 'This sentence is false' represents the quoted sentence itself.) But the normal case is where one thing – the representation itself – represents another thing – what we might call the *object* of representation. We can therefore ask two questions: one about the nature of representations and one about the nature of objects of representation.

What sorts of things can be representations? I have already mentioned words and pictures, which are perhaps the most obvious examples. But, of course, there are many other kinds. The diagram of the hydrogen atom on Pioneer 10's plate is neither a bunch of words nor a picture, but it represents the hydrogen atom. Numerals, such as 15, 23, 1001, etc., represent numbers. Numerals can represent other things too: for example, a numeral can represent an object's length (in metres or in feet) and a triple of numerals can represent a particular shade of colour by representing its degree of hue, saturation and brightness. The data structures in a computer can represent text or numbers or images. The rings of a tree can represent its age. A flag can represent a nation. A political demonstration can represent aggression. A piece of music can represent a mood of unbearable melancholy. Flowers can represent grief. A glance or a facial expression can represent irritation. And, as we shall see, a state of mind – a belief, a hope, a desire or a wish – can represent almost anything at all.

There are so many kinds of things that can be representations that it would take more than one book to discuss them all. And, of course, I shall not try to do this. I shall focus on simple examples of representation in language and in thought. For instance, I will talk about how it is that I can use a word to represent a particular person, or how I can think (say) about a dog. I'll focus on these simple

## The puzzle of representation

examples because the philosophical problems about representation arise even in the simplest cases. Introducing the more complex cases – such as how a piece of music can represent a mood – will at this stage only make the issue more difficult and mind-boggling than it is already. But to ignore these complex cases does not mean that I think they are unimportant or uninteresting.<sup>9</sup>

Now to our second question: what sorts of things can be objects of representation? The answer is, obviously, almost anything. Words and pictures can represent a physical object, such as a person or a house. They can represent a feature or property of a physical object, for example the shape of a person or the colour of a house. Sentences, like the sentence ‘Someone is in my house’, can represent what we might call facts, situations or states of affairs: in this case, the fact that someone is in my house. Non-physical objects can be represented too: if there are numbers, they are plainly not physical objects (where in the physical world is the number 3?). Representations – such as words, pictures, music and facial expressions – can represent moods, feelings and emotions. And representations can represent things that do not exist. I can think about – that is, represent – unicorns, dragons and the greatest prime number. None of these things exist; but they can all be ‘objects’ of representation.

This last example indicates one curious feature of representation. On the face of it, the expression ‘X represents Y’ suggests that representation is a *relation* between two things. But a relation between two things normally implies that those two things exist. Take the relation of *kissing*: if I kiss Santa Claus, then Santa Claus and I must both exist. And the fact that Santa Claus does not exist explains why I cannot kiss him.

But this isn’t true of representation: if I think about Santa Claus, and therefore represent him, it doesn’t follow that Santa Claus exists. The non-existence of Santa Claus is no obstacle to my representing him, as it was to my kissing him. In this way, representation seems very different from other relations. As we shall see later on, many philosophers have taken this aspect of representation to be central to its nature.

## The puzzle of representation

So there are many kinds of representations, and many kinds of things which can be the objects of representation. How can we make any progress in understanding representation? There are two sorts of question we can ask:

First, we can ask *how* some particular kind of representation – pictures, words or whatever – manages to represent. What we want to know is what it *is* about this kind of representation that makes it play its representing role. (As an illustration, I consider below the idea that pictures might represent things by *resembling* them.) Obviously, we will not assume that the story told about one form of representation will necessarily apply to all other forms: the way that pictures represent will not be the same as the way that music represents, for example.

Second, we can ask whether some particular form of representation is more *basic* or *fundamental* than the others. That is, can we explain certain kinds of representation in terms of other kinds. For example: an issue in current philosophy is whether we can explain the way language represents in terms of the representational powers of states of mind, or whether we need to explain mental representation in terms of language. If there is one kind of representation that is more fundamental than the other kinds, then we are clearly on our way to understanding representation as a whole.

My own view is that mental representation – the representation of the world by states of mind – is the most fundamental form of representation. To see how this might be a reasonable view, we need to look briefly at pictorial and linguistic representation.

### Pictures and resemblance

On the face of it, the way that pictures represent seems to be more straightforward than other forms of representation. For, while there is nothing intrinsic to the word ‘dog’ that makes it represent dogs, surely there is something intrinsic to a picture of a dog that makes it represent a dog – that is, *what the picture looks like*. Pictures of dogs look something like dogs – they resemble dogs in some way, and they do so because of their intrinsic features: their shape, colour

## The puzzle of representation

and so on. Perhaps, then, a picture represents what it does because it resembles that thing.

The idea that a picture represents by resembling would be an answer to the first kind of question mentioned above: how does a particular kind of representation manage to represent? The answer is: pictures represent things by resembling those things. (This answer could then be used as a basis for an answer to the second question: the suggestion will be that all other forms of representation can be explained in terms of pictorial representation. But as we shall see below, this idea is hopeless.) Let's call this idea the 'resemblance theory of pictorial representation', or the 'resemblance theory' for short. To discuss the resemblance theory more precisely, we need a little basic philosophical terminology.

Philosophers distinguish between two ways in which the truth of one claim can depend on the truth of another. They call these two ways 'necessary' and 'sufficient' conditions. To say that a particular claim, A, is a *necessary* condition for some other claim, B, is to say this: B is true only if A is true too. Intuitively, B will not be true without A being true, so the truth of A is *necessary* (i.e. needed, required) for the truth of B.

To say that A is a *sufficient* condition for B is to say this: if A is true, then B is true too. Intuitively, the truth of A ensures the truth of B – or, in other words, the truth of A *suffices* for the truth of B. To say that A is a *necessary and sufficient* condition for the truth of B is to say this: if A is true, B is true, *and* if B is true, A is true. (This is sometimes expressed as 'A is true if and only if B is true', and 'if and only if' is sometimes abbreviated to 'iff'.)

Let's illustrate this distinction with an example. If I am in London, then I am in England. So being in England is a *necessary condition* for being in London: I just can't be in London without being in England. Likewise, being in London is a *sufficient condition* for being in England: being in London will suffice for being in England. But being in London is clearly not a necessary condition for being in England, as there are many ways one can be in England without being in London. For the same reason, being in England is not a sufficient condition for being in London.

## The puzzle of representation

The resemblance theory takes pictorial representation to depend on the resemblance between the picture and what it represents. Let's express this dependence more precisely in terms of necessary and sufficient conditions: a picture (call it P) represents something (call it X) if and only if P resembles X. That is, a resemblance between P and X is both necessary and sufficient for P to represent X.

This way of putting the resemblance theory is certainly more precise than our initial vague formulation. But, unfortunately, expressing it in this more precise way only shows its problems. Let's take the idea that resemblance might be a sufficient condition for pictorial representation first.

To say that resemblance is sufficient for representation is to say this: if X resembles Y, then X represents Y. The first thing that should strike us is that 'resembles' is somewhat vague. For, in one sense, almost everything resembles everything else. This is the sense in which resembling something is just having some feature in common with that thing. So, in this sense, not only do I resemble my father and my mother, because I look like them, but I also resemble my desk – my desk and I are both physical objects – and the number 3 – the number 3 and I are both objects of one kind or another. But I am not a representation of any of these things.

Perhaps we need to narrow down the ways or respects in which something resembles something else if we want resemblance to be the basis of representation. But notice that it does not help if we say that, if X resembles Y *in some respect*, then X represents Y. For I resemble my father in certain respects – say, character traits – but this does not make me a representation of him. And, obviously, we do not want to add that X must resemble Y in those respects in which X *represents* Y, as this would make the resemblance theory circular and uninformative: if X resembles Y in those respects in which X represents Y, then X represents Y. This may be true, but it can hardly be an analysis of the notion of representation.

There is a further problem with resemblance as a sufficient condition. Suppose we specify certain respects in which something resembles something else: a picture of Napoleon, for example, might resemble Napoleon in the facial expression, the proportions of the



## The puzzle of representation

body, the characteristic position of the arm, and so on. But it seems to be an obvious fact about resemblance that, if X resembles Y, then Y resembles X. (Philosophers put this by saying that resemblance is a *symmetrical* relation.) If I resemble my father in certain respects, then my father resembles me in certain respects. But this doesn't carry over to representation. If the picture resembles Napoleon, then Napoleon resembles the picture. But Napoleon does not represent the picture. So resemblance cannot be sufficient for pictorial representation if we are to avoid making every pictured object itself a pictorial representation of its picture.

Finally, we should consider the obvious fact that everything resembles itself. (Philosophers put this by saying that resemblance is a *reflexive* relation.) If resemblance is supposed to be a sufficient condition for representation, then it follows that everything represents itself. But this is absurd. We should not be happy with a theory of pictorial representation that turns *everything* into a picture of itself. This completely trivialises the idea of pictorial representation.

So the idea that resemblance might be a sufficient condition of pictorial representation is hopeless.<sup>10</sup> Does this mean that the resemblance theory fails? Not yet: for the resemblance theory could say that, although resemblance is not a sufficient condition, it is a necessary condition. That is, if a picture P represents X, then P will resemble X in certain respects – though not vice versa. What should we make of this suggestion?

On the face of it, it seems very plausible. If a portrait represents the Queen, then surely it must resemble her in some respect. After all, that may be what it is for a portrait to be a 'good likeness'. But there are problems with this idea too. For a picture can certainly represent something without resembling it very much. A lot of twentieth-century art is representational; but this is not to say that it is based on resemblance (consider cubist pictures). Caricatures and schematic drawings, like stick figures, often have very little resemblance in common with the things they represent. Yet we often have no trouble in recognising what it is they represent. A caricature of the Queen may resemble her a lot less than a detailed drawing of someone else. Yet the caricature is still a picture of the Queen.<sup>11</sup>

## The puzzle of representation

So how much resemblance is needed for the necessary condition of representation to be met? Perhaps it could be answered that all that is needed is that there is *some* resemblance, however loose, between the picture and what it represents. Perhaps resemblance can be taken loosely enough to incorporate the representation involved in cubist pictures. This is fine; but now the idea of resemblance is not doing as much work in the theory as it previously was. If a schematic picture (say, of the sort used by certain corporations in their logos) need resemble the thing it represents only in a very minimal way, then it is hard to see how much is explained by saying that 'if a picture represents X, it must resemble X'. So even when a picture does resemble what it represents, there must be factors other than resemblance which enter into the representation and make it possible.

I am not denying that pictures often do resemble what they represent. Obviously they do, and this may be part of what makes them pictures at all (as opposed to sentences, graphs or diagrams). All I am questioning is whether the idea of resemblance can *explain* very much about how pictures represent. The idea that resemblance is a necessary condition of pictorial representation may well be true; but the question is '*What else* makes a picture represent what it does?'<sup>12</sup>

One point that needs to be emphasised here is that pictures often need interpretation. For example, in Michelangelo's *The Last Judgment*, in the Sistine Chapel, we see the souls in hell struggling in agony as they meet their final end, with the monumental figure of Christ above them raising his hand in judgement. Why don't we see the souls being welcomed out of the depths by the benevolent Christ, with his hand raised in friendly encouragement – 'hey, come on up, it's cooler here'? (Remember the picture on Pioneer 10's metal plate of the hand raised in greeting.) Well, we could; but we don't. The reason is that we see the picture in the light of certain assumptions we make about it – what we could vaguely call the 'context' of the picture. We know that the picture is a picture of the last judgement, and that in the last judgement some souls were sentenced to eternal damnation, with Christ as the judge, and so on. This is part of why we see the picture in the way we do: we interpret it.

## The puzzle of representation

We can make the point with an example of Wittgenstein's.<sup>13</sup> Imagine a drawing of a man with a stick walking up a slope (see Figure 1.1). What makes this a picture of a man walking up a slope, rather than a man sliding gently down a slope? Nothing in the picture. It is because of what we are used to in our everyday experience, and the sort of context in which we are used to seeing such pictures, that we see the picture one way rather than another. We have to interpret the picture in the light of this context – the picture does not interpret itself.

I am not going to pursue the resemblance theory or the interpretation of pictures any further. I mention it here to illustrate how little the idea of resemblance tells us about pictorial representation. What I want to do now is to briefly consider the second question I raised at the end of the last section, and apply it to pictorial representation. We could put the question like this: suppose that we had a complete theory of pictorial representation. Would it then be possible for all other forms of representation to be explained in terms of pictorial representation?

The answer to this is 'No', for a number of reasons. One reason we have already glanced at: pictures often need to be interpreted, and it won't help to say that the interpretation should be another picture,

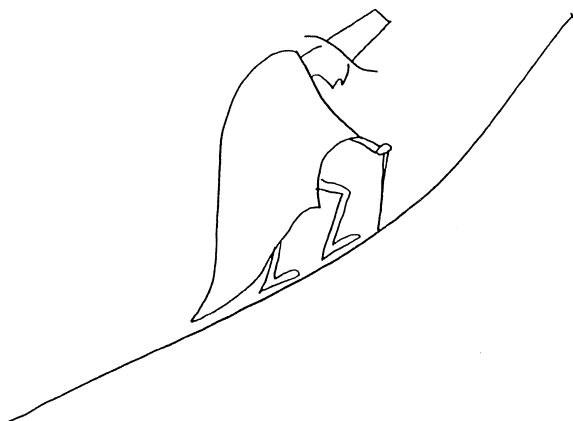


Figure 1.1 Old man with a stick.

## The puzzle of representation

because that might need interpreting too. But, although the answer is 'No', we can learn something about the nature of representation by learning about the limitations of pictorial representation.

A simple example can illustrate the point. Suppose I say to you 'If it doesn't rain this afternoon, we will go for a walk'. This is a fairly simple sentence – a linguistic representation. But suppose we want to explain *all* representation in terms of pictorial representation; we would need to be able to express this linguistic representation in terms of pictures. How could we do this?

Well, perhaps we could draw a picture of a non-rainy scene with you and me walking in it. But how do we picture the idea of 'this afternoon'? We can't put a clock in the picture: remember, we are trying to reduce all representation to pictures, and a clock does not represent the time by picturing it. (The idea of 'picturing' time, in fact, makes little sense.)

And there is a further reason why this first picture cannot be right: it is just a picture of you and me walking in a rain-free area. What we wanted to express was a particular combination and relationship between two ideas: first, it's *not* raining, and, second, you and me going for a walk. So perhaps we should draw two pictures: one of the rain-free scene and one of you and me walking. But this can't be right either: for how can this pair of pictures express the idea that *if* it doesn't rain, *then* we will go for a walk? Why shouldn't the two pictures be taken as simply representing a non-rainy scene *and* you and me going for a walk? Or why doesn't it represent the idea that *either* we will go for a walk *or* it won't rain? When we try to represent the difference between . . . *and* . . . , *if* . . . *then* . . . , and *either* . . . *or* . . . in pictures, we draw a complete blank. There just seems no way of doing it.

One important thing that pictures cannot do, then, is represent certain sorts of relations between ideas. They cannot represent, for example, those relations which we express using the words *if* . . . *then* . . . , . . . *and* . . . , *either* . . . *or* and *not*. (Why *not*? Well, the picture of the non-rainy scene may equally be a picture of a sunny scene – how can we pictorially express the idea that the scene is a scene where there is *no* rain? Perhaps by drawing rain and putting a

## The puzzle of representation

cross through it – as in a ‘No Smoking’ sign – but again we are using something that is not a picture: the cross.) For this reason at least, it is impossible to explain or reduce other forms of representation to pictorial representation.

### Linguistic representation

A picture may sometimes be worth a thousand words, but a thousand pictures cannot represent some of the things we can represent using words and sentences. So how can we represent things using words and sentences?

A natural idea is this: ‘words don’t represent things in any natural way; rather, they represent by *convention*. There is a convention among speakers of a language that the words they use will mean the same thing to one another; when speakers agree or converge in their conventions, they will succeed in communicating; when they don’t, they won’t.’<sup>14</sup>

It is hard to deny that what words represent is at least partly a matter of convention. But what is the convention, exactly? Consider the English word ‘dog’. Is the idea that there is a convention among English speakers to use the word ‘dog’ to represent dogs, and only dogs (so long as they are intending to speak literally, and to speak the truth)? If so, then it is hard to see how the convention can *explain* representation, as we stated the convention as a ‘convention to use the word “dog” to *represent* dogs’. As the convention is stated by using the idea of representation, it takes it for granted: it cannot explain it. (Again, my point is not that convention is not involved in linguistic representation; the question is rather what the appeal to convention can explain on its own.)

An equally natural thought is that words represent by being conventionally linked to the *ideas* that thinkers intend to express by using those words. The word ‘dog’ expresses the idea of a dog, by means of a convention that links the word to the idea. This theory has a distinguished philosophical history: something like it goes back at least as far as Thomas Hobbes (1588–1679), and especially to John Locke (1632–1704), who summed up the view by saying that words are the ‘sensible marks of ideas’.<sup>15</sup>

## The puzzle of representation

What are ideas? Some philosophers have held that they are something like mental images, pictures in the mind. So when I use the word 'dog', this is correlated with a mental image in my mind of a dog. A convention associates the word 'dog' with the idea in my mind, and it is in virtue of this association that the word represents dogs.

There are many problems with this theory. For one thing, is the image in my mind an image of a particular dog, say Fido? But, if so, why suppose that the word 'dog' means *dog*, rather than *Fido*? In addition, it is hard to imagine what an image of 'dogness' in general would be like.<sup>16</sup> And even if the mental image theory of ideas can in some way account for this problem, it will encounter the problem mentioned at the end of the last section. Although many words can be associated with mental images, many can't: this was the problem that we had in trying to explain *and*, *or*, *not* and *if* in terms of pictures.

However, perhaps not all ideas are mental images – often we think in words, for example, and not in pictures at all. If so, the criticisms in the last two paragraphs miss the mark. So let's put to one side the theory that ideas are mental images, and let's just consider the claim that words represent by expressing ideas – whatever ideas may turn out to be.

This theory does not appeal to a 'convention to *represent* dogs', so it is not vulnerable to the same criticism as the previous theory. But it cannot, of course, explain representation, because it appeals to ideas, and what are ideas but another form of representation? A dog-idea represents dogs just as much as the word 'dog' does; so we are in effect appealing to one kind of representation (the idea) to explain another kind (the word). This is fine, but if we want to explain representation in general then we also need to explain how *ideas* represent.

Perhaps you will think that this is asking too much. Perhaps we do not need to explain how ideas represent. If we explain how words represent by associating them with ideas, and explain too how pictures are interpreted in terms of the ideas that people associate with them in their minds, perhaps we can stop there. After all, we can't

## The puzzle of representation

explain everything: we have to take something for granted. So why not take the representational powers of ideas for granted?

I think this is unsatisfactory. If we are content to take the representational powers of the mind for granted, then why not step back and take the representational powers of language for granted? For it's not as if the mind is better understood than language – in fact, in philosophy, the reverse is probably true. Ideas, thoughts and mental phenomena generally seem even more mysterious than words and pictures. So, if anything, this should suggest that we should explain ideas in terms of language, rather than vice versa. But I don't think we can do this. So we need to explain the representational nature of ideas.

Before moving on to discuss ideas and mental representation, I should be very clear about what I am saying about linguistic representation. I am not saying that the notions I mentioned – of convention, or of words expressing ideas – are the only options for a theory of language. Not at all. I introduced them only as illustrations of how a theory of linguistic representation will need, ultimately, to appeal to a theory of mental representation. Some theories of language will deny this, but I shall ignore those theories here.<sup>17</sup>

The upshot of this discussion is that words, like pictures, do not represent in themselves ('intrinsically'). They need interpreting – they need an interpretation assigned to them in some way. But how can we explain this? The natural answer, I think, is that interpretation is something which the *mind* bestows upon words. Words and pictures gain the interpretations they do, and therefore represent what they do, because of the states of mind of those who use them. But these states of mind are representational too. So to understand linguistic and pictorial representation fully, we have to understand mental representation.

### Mental representation

So how does the mind represent anything? Let's make this question a little easier to handle by asking how individual *states* of mind represent anything. By a 'state of mind', or 'mental state', here I mean

## The puzzle of representation

something like a belief, a desire, a hope, a wish, a fear, a hunch, an expectation, an intention, a perception and so on. I think that all of these are states of mind which represent the world in some way. This will need a little explaining.

When I say that hopes, beliefs, desires and so on represent the world, I mean that every hope, belief or desire is *directed at* something. If you hope, you must hope for *something*; if you believe, you must believe *something*; if you desire, you must desire *something*. It does not make sense to suppose that a person could simply hope, without hoping for anything; believe, without believing anything; or desire, without desiring anything. What you believe or desire is what is represented by your belief or desire.

We will need a convenient general term for states of mind which represent the world, or an aspect of the world. I shall use the term 'thought', as it seems the most general and neutral term belonging to the everyday mental vocabulary. From now on in this book, I will use the term 'thought' to refer to all representational mental states. So states of belief, desire, hope, love and so on are all thoughts in my sense, as they all represent things. (Whether all mental states are thoughts in this sense is a question I shall leave until the end of the chapter.)

What can we say in general about how thoughts represent? I shall start with thoughts which are of particular philosophical interest: those thoughts which represent (or are about) *situations*. When I hope that there will be bouillabaisse on the menu at my favourite restaurant tonight, I am thinking about a number of things: bouillabaisse, the menu, my favourite restaurant, tonight. But I am not just thinking about these things in a random or disconnected way: I am thinking about a certain possible fact or *situation*: the situation in which bouillabaisse is on the menu at my favourite restaurant tonight. It is a harmless variant on this to say that my state of hope *represents* this situation.

However, consider a different thought I might have: the *belief* that there is bouillabaisse on the menu tonight. This mental state does not represent the situation in quite the same sense in which the hope does. When I believe that there is bouillabaisse on the menu



## The puzzle of representation

tonight (perhaps because I have walked past the restaurant and read the menu), I take the situation in question to be the case: I take it as a fact about the world that there is bouillabaisse on the menu tonight. But, when I hope, I do not take it to be a fact about the world; rather, I would like it to be a fact that there is bouillabaisse on the menu tonight.

So there are two aspects to these thoughts: there is the 'situation' represented and there is what we could call (for want of a better word) the *attitude* which we take to the situation. The idea of different attitudes to situations is best illustrated by examples.

Consider the situation in which I visit Budapest. I can expect that I will visit Budapest; I can hope that I will visit Budapest; and I can believe that I have visited Budapest. All these thoughts are about, or represent, the same situation – me visiting Budapest – but the attitudes taken to this situation are very different. The question therefore arises over what makes these different attitudes different; but for the moment I am only concerned to distinguish the situation represented from the attitude taken to it.

Just as the same situation can be subject to different attitudes, so the same kind of attitude can be concerned with many different situations. I actually believe that I will visit Budapest soon, and I also believe that my favourite restaurant does not have bouillabaisse on the menu tonight, and I believe countless other things. Beliefs, hopes and thoughts like them can therefore be uniquely picked out by specifying:

- (a) the attitude in question (belief, hope, expectation etc.);
- (b) the situation represented.

(It should also be noted in passing that many attitudes come in degrees: one can want something more or less strongly; and believe something with more or less conviction; but this complication does not affect the general picture.) In general, we can describe these kinds of thoughts schematically as follows. Where 'A' stands for the person who is in the mental state, ' $\psi$ ' stands for the attitude (the Greek letter psi – for 'psychological') and 'S' stands for the situation represented, the best description will be of the following form:

## The puzzle of representation

### A $\psi$ s that S

For example, Vladimir (A) believes ( $\psi$ s) that it is raining (S); Renata (A) hopes ( $\psi$ s) that she will visit Romania (S) – and so on.

Bertrand Russell (1872–1970) called thoughts that can be picked out in this way ‘propositional attitudes’ – and the label has stuck.<sup>18</sup> Though it might seem rather obscure at first glance, the term ‘propositional attitude’ describes the structure of these mental states quite well. I have already explained the term ‘attitude’. What Russell meant by ‘proposition’ is something like what I am calling ‘situation’: it is what you have your attitude towards (so a proposition in this sense is not a piece of language). A propositional attitude is therefore any mental state which can be described in the ‘A  $\psi$ s that S’ style.

Another piece of terminology that has been almost universally adopted is the term ‘content’, used where Russell used ‘proposition’. According to this terminology, when I believe that there is beer in the fridge, the *content* of my belief is that *there is beer in the fridge*. And likewise with desires, hopes and so on – these are different attitudes, but they all have ‘content’. What exactly ‘content’ is, and what it is for a mental state to have ‘content’ (or ‘representational content’), are questions that will recur throughout the rest of this book – especially in Chapter 5. In current philosophy, the problem of mental representation is often expressed as: ‘What is it for a mental state to have content?’. For the time being, we can think of the content of a mental state as what distinguishes states involving the same attitude from one another. Different beliefs are distinguished from one another (or, in philosophical terminology, ‘individuated’) by their different contents. So are desires; and so on with all the attitudes.

I have concentrated on the idea of a propositional attitude, because thoughts of this form will become quite important in the next chapter. But although all propositional attitudes are thoughts (by definition) it is important to stress that not all thoughts (in my sense) are propositional attitudes – that is, not all representational mental states can be characterised in terms of attitudes to situations. Take love, for instance. Love is a representational mental state: you cannot love without loving something or someone. But love is not

## The puzzle of representation

(always) an attitude to a situation – love can be an attitude to a person, a place or a thing. Love cannot be described in the ‘A  $\psi$ s that S’ style (try it and see). In my terminology then, love is a kind of thought, but not a propositional attitude.<sup>19</sup>

Another interesting example is desire. Is this an attitude to a situation? On the face of it, it isn’t. Suppose I desire a cup of coffee: my desire is for a thing, a cup of coffee, not for any situation. On the surface, then, desire resembles love. But many philosophers think that this is misleading, and that it under-describes a desire to treat it as an attitude to a thing. The reason is that a more accurate description of the desire is that it is a desire that a certain situation obtains: the situation in which *I have a cup of coffee*. All desires, it is claimed, are really desires *that so-and-so* – where ‘so-and-so’ is a specification of a situation. Desire, unlike love, is a propositional attitude.

Now, by calling representational mental states ‘thoughts’ I do not mean to imply that these states are necessarily conscious. Suppose Oedipus really does desire to kill his father and marry his mother. Then, by the criterion outlined above (A  $\psi$ s that S), these desires count as propositional attitudes and therefore thoughts. But they are not conscious thoughts.

It might seem strange to distinguish between thought and consciousness in this way. To justify the distinction, we need a brief preliminary digression into the murky topic of consciousness; a full treatment of this subject will have to wait until Chapter 6.

### Thought and consciousness

Consciousness is what makes our waking lives seem the way they do, and is arguably the ultimate source of all value in the world: ‘without this inner illumination’, Einstein said to the philosopher Hebert Feigl, ‘the universe would be nothing but a heap of dirt’.<sup>20</sup> But, despite the importance of consciousness, I want to distinguish certain questions about thought from questions about consciousness. To a certain extent, these questions are independent of one another.

## The puzzle of representation

As I say, this may seem a little strange. After all, for many people, the terms ‘thought’ and ‘consciousness’ are practically synonymous. Surely thinking is being aware of the world, being conscious of things in and outside oneself – how then can we understand thought without also understanding consciousness? (Some people even think of the terms ‘conscious’ and ‘mental’ as synonymous – for them the point is even more obvious.)

The reason for distinguishing thought and consciousness is very simple. Many of our thoughts are conscious, but not all of them are. Some of the things we think are unconscious. So, if thought can still be *thought* while not being conscious, then it cannot *in general* be essential to something’s being a thought that it is conscious. It ought therefore to be possible to explain what makes thought what it is without having to explain consciousness.

What do I mean when I say that some thought is unconscious? Simply this: there are things we think, but we are not *aware* that we think them. Let me give a few examples, some more controversial than others.

I would be willing to bet that you think the President of the United States normally wears socks. If I asked you ‘Does the President of the United States normally wear socks?’ I think you would answer ‘Yes’. And what people say is pretty good evidence for what they think: so I would take your answer as good evidence for the fact that you think that the President of the United States normally wears socks. But I would also guess that the words ‘the President of the United States normally wears socks’ had never come before your conscious mind. It’s pretty likely that the issue of the President’s footwear has never *consciously* occurred to you before; you have never been *aware* of thinking it. And yet, when asked, you seem to reveal that you do think it is true. Did you only start thinking this when I asked you? Can it really be right to say that you had no opinion on this matter before I asked you? (‘Hm, that’s an interesting question, I had never had never given this any thought before, I wonder what the answer is . . . .’) Doesn’t it make more sense to say that the unconscious thought was there all along?

This example might seem pretty trivial, so let’s try a more

## The puzzle of representation

significant (and controversial) one. In Plato's dialogue, *Meno*, Socrates is trying to defend his theory that all knowledge is recollection of truths known in the previous life of the soul. To persuade his interlocutor (Meno) of this, Socrates questions one of Meno's slaves about a simple piece of geometry: if the area of a square with sides  $N$  units long is a certain number of units, what is the area of a square with sides  $2 \times N$  units long? Under simple questioning (which does not give anything away) Meno's slave eventually gets the correct answer. The dialogue continues:

*Socrates:* What do you think, Meno? Has he answered with any opinions that were not his own?

*Meno:* No, they were all his.

*Socrates:* Yet he did not know, as we agreed a few minutes ago.

*Meno:* True.

*Socrates:* But these opinions were somewhere in him, were they not?

*Meno:* Yes.<sup>21</sup>

Socrates, then, argues that knowledge is recollection, but this is not the view that interests me here. What interests me is the idea that one can have a kind of 'knowledge' of (say) certain mathematical principles 'somewhere' in one without being explicitly conscious of them. This sort of knowledge can be 'recovered' (to use Socrates's word) and made explicit, but it can also lie within someone's mind without ever being recovered. Knowledge involves thinking of something; it is a kind of thought. So if there can be unconscious knowledge, there can be unconscious thought.

There are some terminological difficulties in talking about 'unconscious thoughts'. For some people, thoughts are episodes in the conscious mind, so they must be conscious by definition. Certainly, many philosophers have thought that consciousness was essential to all mental states, and therefore to thoughts. Descartes was one – to him the idea of an unconscious thought would have been a contradiction in terms. And some today agree with him.<sup>22</sup>

However, I think that these days many more philosophers (and

## The puzzle of representation

non-philosophers too) are prepared to take very seriously the idea of an unconscious thought. One influence here is Freud's contribution to the modern conception of the mind. Freud recognised that many of the things that we do cannot be fully accounted for by our conscious minds. What does account for these actions are our *unconscious* beliefs and desires, many of which are 'buried' so deep in our minds that we need a certain kind of therapy – psychoanalysis – to dig them out.<sup>23</sup>

Notice that we can accept this Freudian claim without accepting specific details of Freud's theory. We can accept the idea that our actions can often be governed by unconscious beliefs and desires, without accepting many of the ideas (popularly associated with Freud's name) about what these beliefs and desires are, and what causes them – e.g. the Oedipus complex, or 'penis envy'. In fact, the essential idea is very close to our ordinary way of thinking about other people's minds. We all know people whom we think do not 'know their own minds', or who are deceiving themselves about something. But how could they fail to be aware of their own thoughts, if thoughts are essentially conscious?

Anyway, for all these reasons, I think that there are unconscious thoughts, and I also think that we do not need to understand consciousness in order to understand thought. This doesn't mean that I am denying that there is such a thing as conscious thought. The examples I discussed were example of thoughts which were *brought* to consciousness – you brought into your conscious mind the thought that the President of the United States normally wears socks, Meno's slave brought into his conscious mind geometrical knowledge that he didn't realise he had, and patients of psychoanalysis bring into their conscious minds thoughts and feelings that they don't know that they have. And many of the examples I will employ throughout the book will be of conscious thoughts. But what I am interested in is what makes them *thoughts*, not what makes them *conscious*.

In his well-known book, *The Emperor's New Mind*, the mathematician and physicist Roger Penrose claims that 'true intelligence requires consciousness'.<sup>24</sup> It may look as if I'm disagreeing with this remark; but actually I'm not. To say that true intelligence (or

## The puzzle of representation

thought) requires consciousness does not mean that to understand the nature of thought we have to understand the nature of consciousness. It just means that anything that can think must also be conscious. An analogy might help: it may be true that anything that thinks, or is intelligent, must be alive. Maybe. If so, then 'true intelligence requires life'. But that would not *by itself* mean that in order to understand thought we would have to understand life. We would just have to presuppose that the things that think are also alive. Our explanation of thought would not also be an explanation of life. And similarly with consciousness. So I am not disagreeing with Penrose's remark. But I am not agreeing with it either. I am remaining neutral on this question, because I don't know whether there could be a creature that had thoughts, but whose thoughts were wholly unconscious. But, fortunately, I don't need to answer this difficult question in order to pursue the themes of this book.

So much, then, for the idea that many thoughts are unconscious. It is now time to return to the idea of mental representation. What have we learned about mental representation? So far, not much. However, in describing in very general terms the notion of a *thought*, and in articulating the distinction between *attitude* and *content* (or *situation*), we have made a start. We now at least have some basic categories to work with, in posing our question about the nature of mental representation. In the next section I shall link the discussion so far with some important ideas from the philosophical tradition.

### Intentionality

Philosophers have a technical word for the representational nature of states of mind: they call it 'intentionality'. Those mental states which exhibit intentionality – those which represent – are sometimes therefore called 'intentional states'. This terminology can be confusing, especially because not all philosophers use the terms in the same way. But it is necessary to consider the concept of intentionality, as it forms the starting point of most philosophers' attempts to deal with the puzzle of representation.

The term 'intentionality' derives from the scholastic philosophers

## The puzzle of representation

of the Middle Ages, who were very interested in issues about representation. These philosophers used the term ‘intentio’ to mean concept, and the term ‘esse intentionale’ (intentional existence) was used – for example, by St Thomas Aquinas (c.1225–1274) – for the way in which the things can be conceptually represented in the mind. The term ‘intentional existence’ (or ‘inexistence’) was revived by the German philosopher Franz Brentano (1838–1917). In his book *Psychology from an Empirical Standpoint* (1874), Brentano claimed that mental phenomena are characterised:

...by what the scholastics of the Middle Ages referred to as the intentional ... inexistence of the object, and what we, although with not quite unambiguous expressions, would call relation to a content, direction upon an object (which is not here to be understood as a reality) or immanent objectivity.<sup>25</sup>

Things are simpler here than they might initially seem. The phrases ‘intentional inexistence’, ‘relation to a content’ and ‘immanent objectivity’, despite superficial differences between them, are all different ways of expressing the same idea: that mental phenomena involve representation or presentation of the world. ‘Inexistence’ is meant to express the idea that the object of a thought – what the thought is about – exists *in* the act of thinking itself. This is not to say that when I think about my dog there is a dog ‘in’ my mind. Rather, it is just the idea that my dog is *intrinsic* to my thought, in the sense that what makes it the thought that it is is the fact that it has my dog as its object.

I will start by understanding the idea of intentionality as simply as possible – as *directedness on something*. Contemporary philosophers often use the term ‘aboutness’ as a synonym for ‘intentionality’: thoughts have ‘aboutness’ because they are *about* things. (I prefer the term ‘directedness’, for reasons that will emerge in a moment.) The essence of Brentano’s claim is that what distinguishes mental phenomena from physical phenomena is that, whereas all mental phenomena exhibit this directedness, no physical phenomenon exhibits it. This claim, that intentionality is the ‘mark of the mental’, is sometimes called *Brentano’s thesis*.



## The puzzle of representation

Before considering whether Brentano's thesis is true, we need to clear up a couple of possible confusions about the term 'intentionality'. The first is that the word looks as if it might have something to do with the ordinary ideas of *intention*, *intending* and acting *intentionally*. There is obviously a link between the philosophical idea of intentionality and the idea of intention. For one thing, if I intend to perform some action, A, then it is natural to think that I represent A (in some sense) to myself. So intentions may be representational (and therefore 'intentional') states.

But, apart from these connections, there is no substantial philosophical link between the concept of intentionality and the ordinary concept of intention. Intentions in the ordinary sense are intentional states, but most intentional states have little to do with intentions.

The second possible confusion is somewhat more technical. Beginners may wish to move directly to the next section, 'Brentano's thesis' (see p. 36).

This second confusion is between intentionality (in the sense I am using it here) and *intensionality*, a feature of certain logical and linguistic contexts. The words 'intensionality' and 'intentionality' are pronounced in the same way, which adds to the confusion, and leads painstaking authors such as John Searle to specify whether they are talking about 'intentionality-with-a-t' or 'intensionality-with-an-s'.<sup>26</sup> Searle is right: intentionality and intensionality are different things, and it is important to keep them apart in our minds.

To see why, we need to introduce some technical vocabulary from logic and the philosophy of language. A linguistic or logical context (i.e. a part of some language or logical calculus) is intensional when it is non-extensional. An extensional context is one of which the following principles are true:

- (A) the principle of intersubstitutivity of co-referring expressions;
- (B) the principle of existential generalisation.

The titles of these principles look rather formidable, but the logical ideas behind them are fairly simple. Let me explain.

## The puzzle of representation

The principle (A) of intersubstitutivity of co-referring expressions is a rather complicated title for a very simple idea. The idea is just that if an object has two names, *N* and *M*, and you say something true about it using *M*, you cannot turn this truth into a falsehood by replacing *M* with *N*. For example, George Orwell's original name was Eric Arthur Blair (he took the name Orwell from the River Orwell in Suffolk). Because both names refer to the same man, you cannot change the true statement:

George Orwell wrote *Animal Farm*

into a falsehood by substituting the name Eric Arthur Blair for George Orwell. Because the statement:

Eric Arthur Blair wrote *Animal Farm*

is equally true. (Likewise, substituting Eric Arthur Blair for George Orwell cannot turn a falsehood into a truth – e.g. 'George Orwell wrote *War and Peace*.) The idea behind this is very simple: because the person you are talking about is the same in both cases, it doesn't matter to the truth of what you say which words you use to talk about him.

The terms 'George Orwell' and 'Eric Arthur Blair' are 'co-referring terms': that is, they refer to the same object. The principle (A) says that these terms can be substituted for one another without changing the truth or falsehood of the sentence in which they occur. (It is therefore sometimes called the principle of 'substitutivity *salva veritate*' – literally, 'saving truth'.)

What could be simpler? Unfortunately, we don't have to look far for cases in which this simple principle is violated. Consider someone – call him Vladimir – who believes that George Orwell wrote *Animal Farm*, but is ignorant of Orwell's original name. Then the statement:

Vladimir believes that George Orwell wrote *Animal Farm*

is true, while the statement:

Vladimir believes that Eric Arthur Blair wrote *Animal Farm*

## The puzzle of representation

is false. Substitution of co-referring terms does not, in this case, preserve truth. Our apparently obvious principle of the substitutivity of co-referring terms has failed. Yet how can this principle fail? It seemed self-evident.

Why this principle fails in certain cases – notably in sentences about beliefs and certain other mental states – is a main concern of the philosophy of language. However, we need not dwell on the reasons for the failure here; I only want to point it out for the purposes of defining the concept of intensionality. The failure of principle (A) is one of the marks of non-extensionality, or intensionality.

The other mark is the failure of principle (B), ‘existential generalisation’. This principle says that we can infer that something exists from a statement made about it. For example, from the statement:

Orwell wrote *Animal Farm*

we can infer that:

There exists someone who wrote *Animal Farm*.

That is, if the first statement is true, then the second is true too.

Once again, a prominent example of where existential generalisation can fail is statements about beliefs. The statement

Vladimir believes that Santa Claus lives at the North Pole

can be true, while the following statement is no doubt false:

There exists someone whom Vladimir believes lives at the North Pole.

Since the first of these two statements can be true while the second is false, the second cannot logically follow from the first. This is an example of the failure of existential generalisation.

To summarise: intensionality is a feature of sentences and linguistic items; a sentence is intensional when it is non-extensional; it is non-extensional when one or both of the two principles (A) and (B) can fail to apply. Notice that I say the principles *can* fail to apply, not that they must. Of course, there are many cases when we can substitute co-referring expressions in belief sentences; and there are many cases where we can conclude that something exists from

## The puzzle of representation

a belief sentence which is about that thing. But the point is that we have no *guarantee* that these principles will hold for all belief sentences and other 'intensional contexts'.

What has this intensionality got to do with our topic, intentionality? At first sight, there is an obvious connection. The examples that we used of sentences exhibiting intensionality were sentences about beliefs. It is natural to suppose that the principle of substitutivity of co-referring terms breaks down here because whether a belief sentence is true depends not just on the *object represented* by the believer, but on the *way* that the object is represented. Vladimir represents Orwell *as Orwell*, and not *as Blair*. So the intensionality seems to be a result of the nature of the representation involved in a belief. Perhaps, then, the intensionality of belief *sentences* is a consequence of the intentionality of the beliefs themselves.

Likewise with the failure of existential generalisation. The failure of this principle in the case of belief sentences is perhaps a natural consequence of the fact (mentioned above) that representations can represent 'things' that don't exist. The fact that we can think about things that don't exist does seem to be one of the defining characteristics of intentionality. So, once again, perhaps, the intensionality of (for example) belief *sentences* is a consequence of the intentionality of the beliefs themselves.<sup>27</sup>

However, this is as far as we can go in linking the notions of intensionality and intentionality. There are two reasons why we cannot link the two notions further:

- 1 *There can be intensionality without intentionality (representation).* That is, there can be sentences which are intensional but do not have anything to do with mental representation. The best-known examples are sentences involving the notions of *possibility* and *necessity*. To say that something is necessarily so, in this sense, is to say that it could not have been otherwise. From the two true sentences,

Nine is necessarily greater than five

The number of planets is nine

## The puzzle of representation

we cannot infer that:

The number of planets is necessarily greater than five

since it is not necessarily true that there are nine planets. There could have been four planets, or none. So the principle of substitutivity of co-referring terms ('nine' and 'the number of planets') fails – but not because of anything to do with mental representation.<sup>28</sup>

- 2 *There can be descriptions of intentionality which do not exhibit intensionality.* An example is given by sentences of the form 'X sees Y'. Seeing is a case of intentionality, or mental representation. But, if Vladimir sees Orwell, then surely he also sees Blair, and the author of *The Road to Wigan Pier*, and so on. Principle (A) seems to apply to 'X sees Y'. Moreover, if Vladimir sees Orwell, then surely there is someone whom he sees. So principle (B) applies to sentences of the form 'X sees Y'.<sup>29</sup> Not all descriptions of intentionality are intensional; so intensionality in the description is not necessary for intentionality to be described.

This last argument, (2), is actually rather controversial, but we don't really need it in order to distinguish intentionality from intensionality. The first argument will do that for us on its own: in the terminology of necessary and sufficient conditions introduced earlier, we can say that intensionality is not sufficient for intentionality, and it may not even be necessary. That is, since you can have intensionality without any mention of intentionality, intensionality is not sufficient for the presence of intentionality. This is enough to show that these are very different concepts, and that we cannot use intensionality as a criterion of intentionality.<sup>30</sup>

Let's now leave intensionality behind, and return to our main theme: intentionality. Our final task in this chapter is to consider Brentano's thesis that intentionality is the 'mark' of the mental.

### Brentano's thesis

As I remarked earlier, Brentano thought that all and only mental

## The puzzle of representation

phenomena exhibit intentionality. This idea, Brentano's thesis, has been very influential in recent philosophy. But is it true?

Let's divide the question into two sub-questions:

- 1 Do all mental states exhibit intentionality?
- 2 Do only mental states exhibit intentionality?

Again the terminology of necessary and sufficient conditions is useful. The first sub-question may be recast: is mentality sufficient for intentionality? And the second: is mentality necessary for intentionality?

It is tempting to think that the answer to the first sub-question is 'No'. To say that all mental states exhibit intentionality is to say that all mental states are representational. But – this line of thought goes – we can know from introspection that many mental states are not representational. Suppose I have a sharp pain at the base of my spine. This pain is a mental state: it is the sort of state which only a conscious being could be in. But pains do not seem to be representational in the way that thoughts are – pains are just feelings, they are not about or 'directed upon' anything. Another example: suppose that you have a kind of generalised depression or misery. It may be that you are depressed without being able to say what it is that you are depressed about. Isn't this another example of an intentional state without directedness on an object?

Let's take the case of pain first. First, we must be clear about what we mean by saying that pain is a mental state. We sometimes call a pain 'physical' to distinguish it from the 'mental' pain of (say) the loss of a loved one. These are obviously very different kinds of mental state, and it is wrong to think that they have very much in common just because we call them both 'pain'. But this fact doesn't make the pain of (say) a toothache any *less* mental. For pain is a state of consciousness: nothing could have a pain unless it was conscious, and nothing could be conscious unless it had a mind.

Does the existence of sensations refute the first part of Brentano's thesis, that mentality is sufficient for intentionality? Only if it is true that they are wholly lacking in any intentionality. And this does not

## The puzzle of representation

seem to be true.<sup>31</sup> Although we would not say that my back pain is 'about' anything, it does have some representational character in so far as it feels to be in my back. I could have a pain that feels exactly the same, 'pain-wise', but is in the top of my spine rather than the base of my spine. The difference in how the two pains feel would purely be a matter of *where* they are felt to be. To put the point more vividly: I could have two pains, one in each hand, which felt exactly the same, except that one felt to be in my right hand, and the other felt to be in my left hand. This felt location is plausibly a difference in intentionality – in what the mental state is 'directed on' – so it is not true that pains (at least) have no intentionality whatsoever.

Of course, this does not mean that pains are propositional attitudes in Russell's sense. For they are not directed on situations. An ascription of pain – 'Oswaldo feels pain' – does not fit into the 'A  $\psi$ s that S' form that I took as a criterion for the ascription of propositional attitudes. But the fact that a mental state is not a propositional attitude does not mean it is not intentional because, as we have already seen, not all thoughts or intentional states of mind are propositional attitudes (love was our earlier example). And if we understand the idea of 'representational character' or intentionality in the general way that I am doing here, it is hard to deny that pains have representational character.

What about the other example, of undirected depression or misery? Well, of course, there is such a thing as depression in which the person suffering from the depression cannot identify what it is that they are depressed about. But this by itself does not mean that such depression has no object, that it has no directedness. For one thing, it cannot be a criterion for something's being an intentional state that the subject of the state must be able to identify its object – otherwise certain forms of self-deception would be impossible. But, more importantly, the description of this kind of emotion as not directed on anything misdescribes it. For depression of any kind is typically a 'thoroughly negative view of the external world' – in Lewis Wolpert's economical phrase.<sup>32</sup> This is as much true of the depression which is 'not about anything in particular' as of the depression which has a definite, easily identifiable object. The

## The puzzle of representation

generalised depression is a way of experiencing the world *in general* – everything seems bad, nothing is worth doing, the world of the depressed person ‘shrinks’. That is, generalised depression is a way in which one’s mind is directed upon the world – and therefore is intentional – since the world ‘in general’ can still be an object of a state of mind.

It is not obvious, then, that there are any states of mind which are wholly non-intentional. However, there may still be *properties* or *features* of states of mind which are non-intentional: for example, although my toothache does have an intentional directedness upon my tooth, it may have a distinctive quality of *naggingness* which is not intentional at all: the naggingness is not directed on anything, it is just *there*. These apparent properties are sometimes known as *qualia*. If sensations like pain have these properties, then there may be a residual *element* in sensation which is not intentional, even though the sensation considered as a whole mental state is intentional. So even if the first part of Brentano’s thesis is true of whole mental states – they are all intentional – there may still be a non-intentional element in mental life. This would be something of a pyrrhic victory for Brentano’s thesis.

So much, then, for the idea that mentality is sufficient for intentionality. But is mentality necessary for intentionality? That is: is it true that if something exhibits intentionality, then that thing is (or has) a mind? Are minds the only things in the world that have intentionality? This is more tricky. To hold that minds are not the only things that have intentionality, we need to give an example of something that has intentionality but doesn’t have a mind. And it seems that there are plenty of examples. Take books. This book contains many sentences, all of which have meaning, represent things and therefore have intentionality in some sense. But the book doesn’t have a mind.

The natural reply to this is to employ the line of thought I used when discussing linguistic representation above. That is, we should say that the book’s sentences do not have intentionality *intrinsically*, but only have it because they are *interpreted* by the readers of the book. The interpretations provided by the states of mind of the reader, however, do have intrinsic intentionality.



## The puzzle of representation

Philosophers sometimes mark the distinction between books and minds in this respect by talking about ‘original’ and ‘derived’ intentionality. The intentionality present in a book is merely *derived* intentionality: it is derived from the thoughts of those who write and read the book. But our minds have *original* intentionality: their intentionality does not depend on, or derive from, the intentionality of anything else.<sup>33</sup>

So we can reframe our questions as follows: can anything other than minds have original intentionality? This question is very baffling. One problem with it is that if we were to encounter something that exhibited original intentionality, it is hard to see how it could be a *further* question whether that thing had a mind. So do we want to say that only minds, as we know them, can exhibit original intentionality? The difficulty here is that it begins to look like a mere stipulation: if, for example, we discovered that computers were capable of original intentionality, we may well say: How amazing! A computer can have a mind!. Or we may decide to use the terms differently, and say: ‘How amazing! Something can have original intentionality without having a mind!’. The difference between the two reactions may seem largely a matter of terminology. In Chapter 3, I will have more to say about this question.

The second part of Brentano’s thesis – that mentality is a necessary condition of intentionality – introduces some puzzling questions, but it nonetheless seems very plausible in its general outlines. However, we should reserve judgement on it until we discover a little more about what it is to have a mind.

### Conclusion: from representation to the mind

The example of the interstellar ‘letter’ from Pioneer 10 brought the puzzling nature of representation into focus. After that, I considered pictorial representation, and the resemblance theory of pictorial representation, as this kind of representation seemed, at first sight, to be simpler than other kinds. But this appearance was deceptive. Not only does resemblance seem a slim basis on which to found representation, but pictures also need interpretation. Interpretation

## The puzzle of representation

seems necessary for linguistic representation too. And I then suggested that interpretation derives from mental representation, or intentionality. To understand representation, we need to understand representational states of mind. This is the topic of the next chapter.

---

### Further reading

Chapter 1 of Nelson Goodman's *Languages of Art* (Indianapolis, Ind.: Hackett 1976) is an important discussion of pictorial representation. Ian Hacking's *Why Does Language Matter to Philosophy?* (Cambridge: Cambridge University Press 1975) is a very readable semi-historical account of the relation between ideas and linguistic representation. A good introduction to the philosophy of language is Alex Miller's *Philosophy of Language* (London: UCL Press 1997). More advanced is Richard Larson and Gabriel Segal, *Knowledge of Meaning: an Introduction to Semantic Theory* (Cambridge, Mass.: MIT Press 1995) which integrates ideas from recent philosophy of language and linguistics. An excellent collection of essential readings in this area of the philosophy of language is A.W. Moore (ed.) *Meaning and Reference* (Oxford: Oxford University Press 1993). For more on the idea of intentionality, see Chapter 1 of my *Elements of Mind* (Oxford: Oxford University Press 2001). An important discussion is Robert Stalnaker's *Inquiry* (Cambridge, Mass.: MIT Press 1984), Chapters 1 and 2. John Searle's *Intentionality* (Cambridge: Cambridge University Press 1983) is an accessible book on the phenomena of intentionality. A useful collection of essays, many of them quite technical, on the idea of a 'propositional attitude' is Nathan Salmon and Scott Soames (eds.), *Propositions and Attitudes* (Oxford: Oxford University Press 1988). The best one-volume collection of readings in the philosophy of mind in general is still David Rosenthal (ed.), *The Nature of Mind* (Oxford: Oxford University Press 1990). For further reading on consciousness, see Chapter 6 below (pp. 231–232).