

OUP Copyright

CHAPTER 24

Reasons and the content of mental states: 1. reductionist theories

Chapter contents

- Session 1 **Aphasia, deficit studies, modularity, and meaning in cognitive psychiatry** 668
- Session 2 **Preliminaries to a philosophical account of content** 670
- Session 3 **Naturalized or reductionist accounts of content** 673
- Session 4 **Descriptive causal accounts of content** 678
- Session 5 **Teleological causal accounts of content** 680
- Reading guide** 684
- References** 684

OUP Copyright

The central issue of locating meaning in nature

One of the underlying issues facing any science of the mind is to find a place in nature for the meaning, intentionality, or 'content' (as it is called) of mental states. This is certainly a challenge for psychiatry, which embraces at one extreme 'talking cures' from a psychoanalytic tradition and also both drug intervention and psychosurgery. The former deal with the meaningful and symbolic nature of mental states and symptoms. The latter pair both deal with causal effects that have no need for meanings. How can these different perspectives be reconciled? Both this chapter and the next will address the question of how meaningful or 'content-laden' mental states and utterances can be understood as part of the same nature, the same natural world, as causal interactions. (Chapter 26 will address a different but related issue: are reason explanations a species of causal explanation and how is this related to the idea of agency?)

The reason for dividing this topic between two chapters is this. In the *philosophy* of mind there have been two general approaches to accounting for the intentionality of mental states. These are reductionist and non-reductionist accounts. In brief, the former attempt to explain how meaning results from causal processes that can be described in non-intentional terms while the latter deny that any such reduction is possible. (We will return to this distinction shortly.)

These different approaches in the philosophy of mind fit different approaches in *psychiatric* theory and practice. Again very briefly, a reductionist approach to intentionality underpins many cognitivist accounts of both linguistic ability and disability while non-reductionist approaches fit better social or relational or 'discursive psychological' accounts in which the meaning of patients' utterances is constructed in the interaction of patient and clinician. Given these two fundamentally different approaches, the topic of intentionality will be divided between this and the next chapter. This chapter will examine reductionism in the philosophy of mind and cognitivism in psychiatry.

The plan of the chapter

This chapter explores the assumptions which lie behind cognitive approaches to understanding, meaning and thought and looks at philosophical theories which could be pressed into the service of cognitive psychology.

- ◆ *Session 1* examines a standard model drawn from cognitive psychology of both normal understanding of heard speech and related pathologies such as 'pure word deafness'. This model invokes but does not explain the working of a central 'semantic system'. It is this that explains our ability to understand the meaning of words.
- ◆ *Session 2* sets out in a preliminary way the general features of our 'intentionality', which are in need of clarification. 'Intentionality' includes our ability both to think thoughts and to make utterances *about* the world. By setting out its general characteristics we present a standard that any account

of a central 'semantic system' should meet. These characteristics include the *normativity*, *rationality*, and *systematicity* of thought. In setting out the 'explananda', the session looks at Fodor's work, a philosopher whose theories most naturally dovetail with the assumptions of cognitive psychology. (In this chapter rationality will play a minor role. Its importance will be stressed in the contrasting view discussed in chapter 25.)

- ◆ *Session 3* contrasts, again in a preliminary way, two different very general strategies for explaining meaning or intentionality. *Reductionists* such as Fodor strive to explain the workings of a semantic system using concepts that do not presuppose the very properties of meaning or intentionality. (*Antireductionists* are examined in more detail in Chapter 25.) Fodor's version of an internal semantic system is then outlined. It deploys an infinite set of mental representations structured in a 'language of thought'. The account of mental representations outlined so far lacks a key feature, however. This is an account of what gives mental representations (processed in a semantic system) their meaning or *world-involving* nature.
- ◆ *Session 4* presents one solution: a causal theory of reference.
- ◆ *Session 5* presents a different solution: an evolutionary teleological theory. Both solutions are criticized, however, for failing to account for the normativity of thought and language. Finally, the lessons for a future cognitive neuropsychiatry are discussed.

Session 1 Aphasia, deficit studies, modularity, and meaning in cognitive psychiatry

The first reading (linked with Exercise 1—Ellis, 1996) is from a textbook on a cognitivist approach to psychology and psychiatry. Before looking at that, however, it will be worth thinking a little about what cognitive psychology is. Here is a passage from a different recent textbook:

Cognitive psychology can be defined as the branch of psychology which attempts to provide scientific explanation of how the brain carries out complex mental functions such as vision, memory, language and thinking. Cognitive psychology arose at a time when computers were beginning to make a major impact on science and it was perhaps natural that cognitive psychologists should draw an analogy between computers and the human brain. The computer analogy was used frequently to draw up a model of the brain in which mental activity was characterized in terms of the flow of information between different stores.

Parkin (1996, p. 3)

Cognitive psychology, modules, and computers

This passage makes two interesting, and related, claims. One is that cognitive psychology has been very strongly influenced by

developments in computer technology and computer science. The other is that it centres on the idea of information flow between different functional elements of the mind (or the brain?). This latter point presupposes that an analysis of mental capacities can assume a *modular* form. The mind comprises a series of functionally interrelated but also independent modules. As we will see this approach is revealed in the use of flow diagrams connecting modules (such as the 'auditory input lexicon') and showing how the output from one can feed into another.

This presupposition, and the kind of functional analyses it gives rise to, dovetails well with *deficit* studies. By studying mental disabilities, their consequences, and also what abilities are left intact, cognitive psychology can chart underlying functional systems that underpin these abilities. Such studies help to decompose 'normal' abilities into component elements through the examination of abnormal psychopathology.

It is worth noting, however, that there is a substantial assumption at work here. This is the assumption that the divisions that we might draw between abilities correspond to functional divisions between underlying independent (if interrelated) mental (or brain) systems. Now in part this assumption is tested by the future success or otherwise of cognitive psychology, but it is worth noting that there has also been some philosophical debate about whether the mind should be characterized as a system of modules or not. But while reading for that will be flagged at the end, it is not the subject of this chapter.

Talk of 'flow charts' of information flow between modules makes very clear why the computer has proved an influential metaphor. But as we will shortly see, it also encourages a further assumption about how this information flow works which has very important consequences for thinking about how the mind can deal with meanings or mental states with *world-involving* contents.

assumption of cognitive psychology, this might suggest that the chapter would concern two component abilities and thus two underlying functional systems responsible for, (1) recognizing, and (2) understanding words. Now it is worth thinking what sort of account might be given of the latter component. (It is the subject of the bulk of this and the next chapter and there will be time for further reflection.) But in fact, very little is explicitly said in the chapter about *understanding*. The module responsible for that lies at the innermost end of information flow lines (taking the heard word and speech to be at the outer ends) and is the 'semantic system'. Shortly we will return to what is said about its working but first think about the form of analysis offered for word *recognition*. Take this introductory passage:

Imagine the case of recognizing a single word, clearly articulated and spoken in isolation. Unless that word is a homophone its sound pattern will be unique to it. To identify the word a listener will need to have stored in memory all the sound patterns of words he or she knows, and be able to compare the pattern just heard with these stored patterns to find the best match. What we are proposing is... [a] word store or lexicon... the auditory input lexicon. (p. 143)

The computer metaphor is at work in this passage in the assumptions it reveals about the underlying processing involved in word recognition. It says that a speaker (hearer) will need to have all the sound patterns of words he or she knows 'stored up in memory' and that these will then be compared with the pattern just heard to obtain the best match. But who has this memory and who does the comparing? One could imagine going about the laborious business of comparing sounds with, perhaps, a set of recordings of words spoken in a foreign language. But this is not what anyone actually does. Likewise, in order to recognize a word, one will have to remember what it sounds like but that is not to say that the sounds are 'stored' in memory.

This is not really a fair line of criticism. But what it reveals is a dangerous slippage between talk of a person's memory and the analogical talk of computer memory. Assumptions about computational processing underpin these strange locutions about comparing sounds with stored patterns.

Explanation of psychopathologies through the breakdown of functional connections

The chapter (Ellis, 1996) goes on to describe three different routes between hearing and repeating words. The most direct and shallow route with the least processing runs directly from a phonetic breakdown in the 'auditory analysis system' to the phoneme level of the output. This is the route used in the case of repeating nonsense words. The deepest route requires the recognition of words on the basis of phonemes by the auditory input lexicon, the understanding of the word in the semantic system, its activation of the speech output lexicon, and finally, activation of the phoneme level and speech. Psychopathological cases can then be explained through the selective breakdown of some but not all of these routes.

EXERCISE 1

(30 minutes)

Read the extract from:

Ellis, A. (1996). Recognising and understanding spoken words. In *Human Cognitive Neuropsychology: a textbook with readings*. Hove: Psychology Press, chapter 6. (Extract: pp. 143–146)

Link with Reading 24.1

- ◆ What is the relationship between pathologies of understanding and underlying modules?
- ◆ What account of normal successful understanding is given?

Recognizing *versus* understanding

Note first that the chapter title of this reading by Ellis (1996) is 'Recognising and understanding spoken words'. Given the modular



Given these different routes, a distinction between 'pure word blindness' and 'word meaning deafness' is then possible. The former involves 'impaired speech perception in the context of good speech production... and, importantly, intact perception of non-verbal environmental sounds' (p. 147). In such cases, subjects are unable to repeat words spoken to them. By contrast, word meaning deafness does not involve this last inability, although it does involve an inability to understand spoken words combined with an ability to understand written words.

1. The word must have undergone adequate acoustic analysis as evidenced by correct repetition.
2. The semantic representation of the word must be intact as evidenced by immediate comprehension of the word when presented in written form.

Intact repetition of words and sentences implies an intact early stage of auditory analysis (the stage thought to be impaired with pure word deafness). Intact reading comprehension and spontaneous speech imply an intact semantic system and speech output lexicon... [W]ord-meaning deafness represents a complete or partial disconnection of the auditory input lexicon from the semantic system. (pp. 154–155)

Finally, there is the case of 'auditory phonological agnosia' in which subjects cannot repeat invented words but can repeat known real words. This can briefly be described as a breakdown of the direct connection between the auditory analysis system and the phoneme level of the output.

But is there an explanation of normal understanding?

These three types of psychopathological case help to support the particular functional hierarchy proposed. Both the component modules and their interconnections are derived from just such cases. In a sense, the comparison between normal and abnormal cases is an application of Mill's Method of Difference (see chapter 16). Thus the flow chart helps to shed light on what can go wrong with the recognition and understanding of words. But there is another sense in which it does not help shed much light on understanding because it does not break down what is going on in the central semantic system. It is one thing to explain a *failure* of understanding in the case of word meaning deafness by saying that there is a lack of information flow to the semantic system but quite another thing to say how that system underpins semantic understanding in other cases.

There are, however, some clues to the sort of account, which—if successful—would fit with a cognitive psychological approach. Think again about the sort of mention made of the semantic system in the reading (Ellis, 1996):

We propose that the first stage of auditory word recognition performed by an early auditory analysis system attempts to identify phonemes in the speech wave. The results of this analysis are transmitted to the auditory input lexicon where a match

is sought against the stored characteristics of known words. If the match is a good one, the appropriate recognition unit in the auditory input lexicon will be activated. It, in turn, will then activate the representation of the meaning of the heard word in the semantic system... (p. 144)

The key phrase here is: 'the representation of meaning of the heard word in the semantic system'. If there is to be a satisfactory account of how one can understand the meaning of a word on hearing it this comment would have to be unpacked. What is it to represent the meaning of a word? How could a module of the mind or brain represent meanings? Compare that issue with, for example, other modules that supposedly store acoustic properties of words. Is the *meaning* of a word a property akin to its acoustic properties? But it is also worth asking how an auditory analysis system could *represent* acoustic properties. We know how *people* might represent such properties. They might use a scientific vocabulary and notation and write down details using these. But how could a mindless *system* do this?

These are serious questions. But there is a clue to a possible approach to solving them that comes from the comment above that cognitive psychology was influenced by computers. These are machines that manipulate symbols in just such a way that information processing is possible. So perhaps answers to the question of how meaning or representation within a semantic system is possible can be answered by looking at the design of computers. That is the thought we will follow in the next session.

Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What does the reading suggest about the challenge of explaining understanding in cognitive psychiatric terms?

Session 2 Preliminaries to a philosophical account of content

The first reading sets out what needs explanation

This session will begin to think about the general shape of the philosophy of thought, intentionality, or content. Chapter 23 looked at some accounts of the relation of mind and body or brain. But one feature of at least some mental states was largely absent from the discussion. This is that some mental states—e.g. thoughts—can be *about* things. One can think about one's cat Brix or one's next birthday (whether or not in fact one lives to see it). One can hope that one will have a nice party or that Brix will not eat too much cake. How is this possible?



EXERCISE 2

(30 minutes)

The purpose of this exercise is to start to think about what a philosophy of thought, or intentionality, or content has to account for. Later, we will read Fodor's (1935–) positive theory of mental content.

First, write down some of the features or properties of mental states, especially mental states that possess intentionality or aboutness, i.e. thoughts. Which of these features are distinctive of mental as opposed to physical states?

Then read the extract from:

Fodor, J.A. (1991). Propositional attitudes. In *The Nature of Mind* (ed. D. Rosenthal). Oxford: Oxford University Press, pp. 325–338. (Extract: pp 325–326)

Link with Reading 24.2

Make a note of the less contentious, pre-philosophical characteristics of content. How similar are these properties to the properties of non-mental objects? Think also about what a philosophical account of content-laden mental states would have to explain. A central question is: how can mental states be about anything? But are there other important questions?

The general characteristics of 'content'

In philosophy the word 'content' is used to refer to both mental content and linguistic meaning. What these both have in common is the possession of intentionality or 'aboutness'. Content-laden *mental* states—such as beliefs, intentions, and expectations—and *linguistic* entities—such as sentences and utterances—can be *about* distant or even non-existent states of affairs. In ascribing content to mental states or to groups of words, one ascribes this kind of general intentionality or 'aboutness'. In this broad philosophical sense, 'intentionality' refers to the capacity to be about something. This is a philosophical and non-standard use of the word. It should not be confused with the everyday sense of intentional, which means deliberate or with an intention. Intentions in that other sense are merely instances of mental states with the broader property of intentionality. The intention that one will read more philosophy possesses the broader property of intentionality because it concerns, and is about, one's intended future reading practice.

Brentano

There is a long philosophical tradition, attributed to Brentano's work at the turn of the century, which takes intentionality in this abstract general sense—the sense of 'aboutness'—to be the defining feature of the mental. Franz Brentano (1838–1917) was a philosopher and psychologist who repopularized the scholastic word 'intentionality' to characterize the world-relatedness of mental states. His most important work was his 1874 *Psychology from the Empirical Standpoint*.

Despite the close connections between the mind and intentionality, it is worth noting two qualifications of any such definition. First, it should not be taken to imply that the only forms of intentionality are forms of mental content. Sentences and utterances also possess intentionality. Secondly, there are mental states, such as sensations, which at least arguably lack intentionality. ('Arguably' because while it makes no sense to ask someone who has reported that they have a pain what it is about, one can ask where it is and this is something that the sensation itself tells them.) Nevertheless, accounting for the intentionality of mental states is an important philosophical task whether or not this takes the form of a reductionist explanation of intentionality. Brentano himself denied that intentional notions could be reduced to non-intentional terms.

'That-clauses'

A convenient way of spelling out the content of a mental states is the use of a 'that-clause'. The content of the belief that psychotherapy is cost-effective is, precisely, that *psychotherapy is cost-effective*. The content of the hope that the drug treatment will work is that *the drug treatment will work*. Using the content is clearly not the only way of *referring* to mental states. But contents as specified by 'that-clauses' are necessary individuating properties of content-laden mental states. Such specification also unpacks the intentionality that mental states possess. Both the object of the belief or the intentional object—psychotherapy—and what is attributed to it—its cost-effectiveness—are stated in the appropriate 'that-clause'.

As content can be stated using a that-clause it is often also called *propositional* content. Content-laden mental states are propositional attitudes because they can be individuated as attitudes or relations towards propositions. Smith's hope that the drug treatment will work is an attitude of hopefulness that Smith has towards the proposition *the drug treatment will work*. Smith is related via a relation of hope to that proposition. This form of analysis is clearly a move towards a philosophical rather than a pre-philosophical account of content. But it is a more or less natural development from and codification of the claim that content is what is stated by a 'that' clause.

Normativity

Spelling out intentionality in terms of propositional content highlights a further property or characteristic of intentional mental states or linguistic items. Propositional content is normative. This is clearest in the case of beliefs and assertions. A belief or assertion is either true or false in virtue of its content. It is true if its subject matter is as it takes it to be and false otherwise. It *prescribes* the condition that has to obtain for it to be true. However, other content-laden mental states are normative in that they prescribe what would satisfy them. Propositional attitudes specify a condition for the world to satisfy if the attitude is to be fulfilled. Linguistic meaning shares this normativity. Sentences and utterances also prescribe their truth conditions.

Fodor's account of mental states

Fodor's three criteria and the role of folk psychology

As we have seen, Fodor (1991) has his own list of essential properties of content-laden attitudes. He gives the following, different list in his *Psychosemantics* (1987, p. 10). These are:

1. Content-laden mental states are *semantically evaluable*. Fodor explains this by saying that they are the kinds of things that are true or false (beliefs), are fulfilled or frustrated (desires), turn out to be right or wrong (hunches), and so on. This is to emphasize the fact that they can be expressed by that-clauses and are normative. It is, in other words, an implicit part of what is ordinarily understood by content-laden mental states.
2. Fodor claims that propositional attitudes have *causal* powers. This is a more contentious claim. Indeed a later chapter will be given over exclusively to discussing it (Chapter 26). Whether or not mental states have causal properties will also be discussed as part of the discussion of reductionist theories in this chapter. But for now, it will be put to one side.
3. Mental states are governed by a common-sense belief desire psychology. Propositional attitude psychology is our everyday way of making sense of, explaining and predicting other people's behaviour. We explain the actions of others by attributing to them beliefs, desires, hopes, fears, and intentions. These can be used to make sense of or rationalize their actions. Because this method of action explanation is used in this everyday way it is often called folk psychology. It works by deploying a complex *system* of propositional attitudes to make sense of action and speech.

In his book *Mental Content* (1989), the philosopher of mind Colin McGinn (1950–) characterizes the connection between mental states and folk psychology in this way:

First reference is made to a particular *person*; then some *attitude* is ascribed to that person; finally a *content* is specified for that attitude. Assertions of this form tell us who has what attitude towards which proposition. By making and receiving such assertions we come (it seems) to understand other people: what they do, why they want such and such, what made them hope for so and so, and so forth. Varying the three elements in the *person-attitude-content* structure gives us a seemingly powerful system for describing the minds of others (and our own), a system both antique and ubiquitous. Thinking of this system as a (tacit and unformalised) theory, we can say that folk psychology is a theory that centrally employs an explanatory ontology of persons and contentful attitudes; with these basic theoretical resources it sets about its explanatory and descriptive work.

McGinn (1989, p. 120)

(Linguistic meaning dovetails with this system in the following way. Utterances provide the most important evidence for the ascription of propositional attitudes because people can *express* their mental states through language. The content of mental states can be put into words without remainder. There is

considerable philosophical dispute—based in part on a difference of pre-philosophical intuitions—about whether mental states also possess qualitative characteristics that are ineffable. Wittgenstein, for example, provides arguments against such a possibility. What matters in this context, however, is the *content* of beliefs and the claim that that is fully expressible is less controversial.)

Is folk psychology a theory?

On the assumption that folk psychology can be characterized as a *theory*, there has been much recent philosophical debate about whether it is a good theory. We will return to the question of its theoretical character in a later chapter. For now it is enough to note that folk psychology is a *systematic* interpretative stance, that it is used to explain behaviour and that it does this by ascribing content-laden mental states. (This is again a subject of Chapter 27 on how we can have access to others' mental states. Do we employ an interpretative theory or do we simulate what they are thinking by imaginatively putting ourselves into their position?)

Mental states possess a characteristic systematic structure

The systematicity of folk psychology just mentioned presupposes some relational structure between the content-laden mental states ascribed. It is worth mentioning two such connections here. One is that content is *rationally structured*. This is a claim forcefully stated by Frege. Mental states stand in logical and rational relations to one another. They are consistent, inconsistent, justificatory, disconfirmatory, and suchlike. As a result, the ascription of mental states in the explanation of speech and action through folk psychology is in part normative and prescriptive. The ascription turns in part on what people ought to think in the circumstances. The extent of this normativity is a philosophical matter, rather than merely a descriptive pre-philosophical one. For now, it suffices to say that the evidence for or against the ascription of an attitude to a person includes the rational connections between that attitude and the other attitudes already ascribed. Thus content-laden states are rationally structured.

There is a second and distinct respect in which content is structured, although this is clearly more of a philosophical claim, albeit one based on a powerful intuition. If one is able to understand the thought that a frog is green, one must also be able to understand thoughts to the effect that things other than frogs are green and that other frogs have colours other than green. Obviously, this is not to say that one must actually *hold* that this is the case. (One may have good reasons to hold that there is only one frog and one green thing left in the world.) But the ability to *understand* the first thought presupposes the ability to understand these other thoughts because understanding is a structured ability. The most influential statement of the claim that understanding a thought or, more generally entertaining a

propositional attitude, is a structured ability that systematically presupposes connections to other thoughts or propositional attitudes is by the Oxford philosopher Gareth Evans (1946–80) in his difficult work *The Varieties of Reference* (1982, pp. 100–105). Consequently, it is often termed Evans's *Generality Constraint*. But the claim that content is systematic in this way is also emphasized by many others including Fodor (1987, pp. 147–153). (Linguistic meaning is also systematic. When one comes to understand a language—rather than simply knowing some phrases from a phrase book—one understands how to construct grammatically correct sentences from constituent words. It is criterial of this sort of understanding that if one can understand the sentence 'John loves Mary' then one will necessarily understand the sentence 'Mary loves John'. When one understands a language, one has a systematic ability to construct and understand novel sentences that employ known words in new structures. As we will see, Fodor uses the systematicity of our understanding of language to attempt to explain the systematicity of our mental states via the postulation of a 'language of thought'.)

The big question

These general characteristics of content-laden mental states present the challenge facing any philosophical explanation of content. How can there be states with just these properties?

Consider the sorts of properties that a chair has and compare this with those that a belief—perhaps a belief about a chair—has. The chair has mass, weight, length, and colour. It might also be to the left of the table or on top of the rug. The belief, by contrast, does not seem to have any of these properties. We would not normally say that a belief had a mass, length, or colour. And although the person having the belief may be to the left of the door or on top of the rug, we would not normally say that that was where the *belief* was. We don't normally say that beliefs are anywhere (except metaphorically perhaps at the back of my mind).

On the other hand, the belief is *about* the chair and it is held or entertained *by* a particular person. These are *relations* between the belief and parts of the world and the person holding it. Suppose the belief is that the chair is yellow. This looks like a relation between a person and the fact that *the chair is yellow*. Now some of the properties listed for the chair itself are also relational. The positional properties are clearly relational, linking the chair with the table and rug. Weight also turns on a gravitational field. It is arguable that even properties such as mass are really relations between different objects. But none of these relations mirror intentional relations. Note, for example, that the belief about the chair might be *false*. (Thus it would not be a relation to the yellowness of the chair: the chair is not yellow.) Or it might even be a belief about a non-existent chair. (The chair at the top of Mount Everest.) By contrast the relational properties of the chair have to link it to real things. So again the big question is: How can there be such things as beliefs with the properties they have?

Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What mark of the mental did Brentano emphasize?
2. What are 'propositional attitudes'?
3. What aspects of mental states does Fodor stress?
4. What is the central question of the philosophy of content?

Session 3 Naturalized or reductionist accounts of content

What sort of account of content is needed?

The previous session attempted to give an everyday description of the features of content-laden mental states, or propositional attitudes. The next reading (linked with Exercise 3—Fodor, 1987) will begin to focus on one sort of philosophical approach to explaining how there can be states with these features. In effect, it provides one approach for augmenting cognitivist psychological theorizing about, for example, word recognition and understanding with an account of how the central 'semantic system' might work. It might thus shed light on how both linguistic understanding and the formation of mental contents (both of which are semantic) are possible.

What is the relation between the Representational Theory of Mind and functionalism?

This session will begin to examine a philosophical account of the meaning or content of mental states, which, like cognitive psychology, embraces the metaphor of the mind as computer. It is a form of cognitivism or representationalism that attempts to explain thought by postulating a system of internal mental representations within a structure of a language of thought. It is the contemporary American philosopher and psychologist Jerry Fodor's Representational Theory of Mind (RTM).

We have already examined one approach to the philosophy of mind based on an analogy with computers. That position was functionalism. (See Chapter 23.) It may be helpful to have an early hint at the difference between functionalism and the position that Fodor sets out, which is the subject of this session: the RTM. There are two key differences. (Do not worry if you do not understand these differences. They are the subject of this session.)

1. Functionalism is an answer to the question: What are mental states and how are they related to states of the body? Functionalism replies that mental states are second order states, individuated by their functional causal properties,

however those are realized. (We saw that some functionalists went on to argue that such states are physical, but that is a further argument.)

Fodor's RTM is an answer to the further question: How can mental states have meaning or content, how can they be *about* worldly states of affairs? A thorough functionalism would also have to answer this question but that is not how that position arose.

2. Functionalism individuates states by their functional properties. Mental states with different contents or meanings would have to be individuated by different functional properties. Thus the belief that coffee is good would have to have a different functional role to the belief that coffee is healthy. A functionalist would have to explain what this different property was without simply saying that one has the functional role of the belief that coffee is good while the other has the functional role that coffee is healthy, and go on to account for all other possible beliefs.

Fodor thinks that such an account will fail to cope with the infinity of potential belief types. He proposes instead to explain different mental contents through the idea of different internal mental representations, which are part of an inner innate *language* of thought. This move to talk of an inner language goes beyond the resources of functionalism proper. (Distinguishing carburettors from exhaust manifolds does not require talk of a *language* of motorcar components.)

The reading is taken from what is perhaps the clearest of Fodor's accounts of his views. Fodor has continued to develop and change his views and so this 1987 account from his short book *Psychosemantics* is not the most up to date, but it will help set out the general approach.

EXERCISE 3

(30 minutes)

Read the extract from:

Fodor, J.A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press, chapter 4. (Extract: pp. 97–99)

Link with Reading 24.3

Think generally about the kind of explanation Fodor offers. What reason does Fodor give for the need for such a theory? What is the relation between explanatory priority of mental content and linguistic meaning?

Reductionist explanations of content

Fodor's account of mental content aims at a form of reductionism. Reductionist accounts aim to explain content by showing how it derives from processes that can be described without using, and thus presupposing, any intentional concepts. The aim is that the puzzling features of content can be reduced to processes which are not puzzling. Fodor characterizes this aim as

follows: 'What we want at a minimum is something of the form "R represents S" is true iff C where the vocabulary in which condition C is couched contains neither intentional nor semantic expressions.' (*Psychosemantics*, p. 32).

Reductionism aims at giving necessary and sufficient conditions (or failing that at least sufficient conditions) for content which can themselves be stated without using intentional concepts. If this could be achieved then it would provide a translation of the supposedly philosophically problematic intentional vocabulary into an unproblematic language and thus resolve the philosophical difficulties. Intentional states would be shown to comprise or derive from non-intentional states. But this raises the question: why think of talk of meaning, intentionality, or content as problematic?

Naturalism

A clue lies in the fact that reductionism here is sometimes called a 'naturalistic' account of content. Such a label reflects the belief that unless content can be explained in non-intentional terms then it cannot be an unmysterious part of the natural world. Fodor gives a clear expression of this methodological assumption in the following passage:

I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin*, *charm* and *charge* will perhaps appear upon their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. It's hard to see... how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. If the semantic and intentional are real properties of things, it must be in virtue of their identity with... properties that are *neither* intentional *nor* semantic. If aboutness is real, it must be really something else. (p. 97)

Fodor's claim is thus that unless one can reduce or 'naturalize' the intentional, then the only other plausible view is to deny that it really exists. The motivation for this claim appears to be that, if it is not reducible to something else, nothing as strange and magical as meaning could itself form part of a respectable account of the world. Failing reduction, one would be forced to adopt an anti-realist reconstrual of content as a strictly fallacious way of speaking, a hangover from the pre-scientific age or as a sort of secondary quality existing only in the eye of the beholder. This is an assumption to which we will return in the next chapter but the connection with cognitive psychology should be clear.

Fodor and cognitive psychology share reductionist and naturalistic assumptions

Cognitive psychology also aims at a reductionist account of human mentality. This is implicit in its attempt to analyse human cognitive abilities in terms of information flow between different subpersonal systems. The very idea of talking of a semantic system as a subsystem of a whole person reveals the assumption that the semantic properties of mental states of whole people can

be explained through lower level information processing. While Fodor makes explicit the motivation for a reductionist explanation of content-laden mental states, this motivation is shared at an implicit level by the assumptions made about cognitive psychological explanation. (This leads to the obvious question: if this motivation is questionable, what other kind of account of semantically evaluable states is possible. One such approach will be the subject of Chapter 25.)

Explanatory priority

A second preliminary observation of Fodor's approach is of the order of explanatory priority, which he assumes. Given that the aim is to provide a reductionist or 'naturalistic' account of intentionality, and given that both mental states but also linguistic items can possess intentionality, there are a number of different options for explanatory priority. What is to be explained in terms of what?

EXERCISE 4

(10 minutes)

Stop! Before looking ahead think about the options available for a reductionist account of intentionality. What approaches might there be? Write down brief notes if you can on these points before going on.

Four explanatory options

There are four possible approaches to explanatory priority:

1. Assign explanatory priority to mental content and explain how linguistic meaning results from that.
2. Assign explanatory priority to linguistic meaning and explain how mental content results from that.
3. Explain mental content and linguistic meaning in the same way with equal priority.
4. Explain mental content and linguistic meaning in different ways with equal priority.

Option 1

The first option is to attempt to explain linguistic meaning by showing how it results from mental content. This is the strategy adopted by Fodor and the majority of reductionist theorists. Their hope is that conditions can be specified concerning the mental states of speakers, which are necessary and sufficient to establish linguistic meaning. The standard approach is to attempt to explain the meaning of sentences as an abstraction from the meaning of utterances made using them. The meaning of utterances is then supposed to be explained as deriving from the content of beliefs that the utterances were intended to convey. Given the content of these beliefs and intentions, the meaning of sentences can be derived and explained. A different and independent account would then have to be found of how mental states possess their content.

Examples of this strategy dominate contemporary philosophy of thought and language. This is partly the result of the influence

of Grice's work in the 1950s (see for example Grice, 1969). Grice aimed to explain linguistic meaning as the result of speakers' beliefs and intentions. He attempted to articulate the conditions that have to be met for a speaker to intend to communicate a belief to an audience using a linguistic expression. The promise of the Gricean programme is that the meaning of conventional spoken words and sentences is explained as deriving from the content of the beliefs that the speaker intends to communicate and the audience realizes it is intended to share by means of established linguistic conventions. Thus linguistic meaning can be explained in terms of a speaker's beliefs and intentions. Attempts based upon Grice's work to reduce in this way the two problems of linguistic meaning and mental content to a single problem of mental content are called 'intentional based semantics'.

Once the two problems are reduced to one, reductionist accounts then attempt to provide a reductionist explanation of mental content in non-intentional terms. But it is worth noting that the question of the order of priority is independent of the question of whether to give a reductionist explanation or not. As we will see, however, the first order of priority fits most easily with reductionism.

Options 2 and 3

There are three other strategies concerning explanatory priority but only two of them are practical. The second is to reverse the above order and to explain mental content in terms of linguistic meaning. The third is to ascribe priority to neither side and to explain both in the same neutral way. While these are clearly different in principle, in practice it is difficult sharply to distinguish examples of them. Philosophers such as Wittgenstein and Davidson appear to subscribe to one or other of these two views but there are conflicting reasons for viewing their approaches as belonging to both the second and third strategy. (They are the subject of the next chapter but the reasons for ascribing to both the second and third position is this. Both Wittgenstein and Davidson provide an argument that the possession of a language is a precondition of any but the most rudimentary mental states. But they also explain linguistic meaning and mental content in the symmetric manner.)

For reasons that will become clearer, the second and third order of priority would be difficult to combine with a reductionist explanation of intentionality. Thus although they tie mental content and linguistic content together, they are not, as a matter of fact, the preliminary stage of a form of reductionism.

Option 4

There is also, in principle, a fourth possibility with respect to explanatory priority. One could assign no priority and offer two different and independent explanations of mental and linguistic content. Such a strategy has two clear practical disadvantages. First, it assumes from the start that the task is twice as difficult as the other three approaches. It assumes that two different accounts of content have to be given rather than that one will serve for

both mind and language. Secondly, it faces the problem of explaining how the two distinct forms of content are related. Given that the content of a belief can be expressed by a sentence, some account of this congruence is required. It is unsurprising that examples of this strategy are difficult to find.

Fodor's representational theory of mind

The Representational Theory of Mind and the language of thought as a reductionist theory

So much then for the kinds of approach to a philosophical account of content. Let us return to Fodor's theory. In order to provide a reductionist explanation of how meaning or content can be part of the natural order, he deploys a causal theory of mental content, which he ties to the Language of Thought hypothesis.

The heart of the Representational Theory of Mind

Fodor suggests that the Representational Theory of Mind (RTM) is defined by two central claims. In *Psychosemantics* he says:

At the heart of the theory is the postulation of a language of thought: an infinite set of 'mental representations' which function both as the immediate objects of propositional attitudes and as the domains of mental processes. More precisely, RTM [the Representationalist Theory of Mind] is the conjunction of two claims:

Claim 1 (the nature of propositional attitudes):

For any organism *O*, and any attitude *A* toward the proposition *P*, there is a ('computational'/'functional') relation *R* and a mental representation *MP* such that

MP means that *P*, and

O has *A* iff *O* bears *R* to *MP*

Claim 2 (the nature of mental processes):

Mental processes are causal sequences of tokenings of mental representations. (pp. 16–17)

What are mental representations?

On a first reading this may seem a rather daunting paragraph; however, it can be understood as a development of a functionalism in the philosophy of mind. Fodor claims that he can explain what it is to have a propositional attitude or content-laden mental state by invoking internal mental representations. As a first approximation, think of mental representations as internal second order states akin to functional states. There are some differences between Fodor's mental representations and functional states. This is signalled here by the phrase 'language of thought'. These differences, however, will become clearer shortly.

Mental representations are inner vehicles of content

Take the first of Fodor's claims. A person, *O*, has a propositional attitude or content laden mental state, *A*, if and only if they stand in a certain relation, *R*, to an internal mental representation *MP*. So Fodor *explains* what it is to have a mental state by hypothesizing

an internal state of the mind or brain, *MP*. The mental representation *encodes* the content of the propositional attitude. Thus it might somehow encode the content that *psychotherapy is cost-effective*. One way of construing encoding is to compare it with a written sentence or phrase. Take the sentence: 'Psychotherapy is cost-effective.' As a sentence it is composed of a number of words that are in turn made up of letters. The whole sentence is a squiggle on paper or a computer screen. But on the assumption that it is a sentence in English it says that *psychotherapy is cost-effective*. Those squiggles 'encode' that content. They are the vehicle of the content. Fodor thinks that in a similar way, mental content is carried by internal vehicles of content, which in this case are mental representations. These encode the contents of our mental states.

Relation R explains the differences between kinds of attitude

What of the relation *R* that mental representations are supposed to stand in? Fodor calls it a functional or computational relation. Here the idea is that the same content can be the object of different attitudes. One may believe, hope or fear that psychotherapy is cost-effective. In each case that content will have a different effect on one's mental economy. One acts differently on the basis of beliefs, hopes, and fears. Fodor suggests that these differences can be explained as the result of mental representations standing in different functional or computational relations.

So the role of mental representations is to explain how it is possible to have content-laden mental states. Part of the account employs functionalism in order to explain and distinguish the different 'psychological' attitudes described in folk psychology. So far, however, nothing has been said about how mental representations encode content. This is what Fodor's version of the causal theory of content is supposed to explain. As we will see, the basic idea of that is that mental representations encode contents by standing in suitable causal relations to those items in the external world which they are about. But before examining that theory, there is a further detail worth examining. This is how the structure of content is explained.

The role of the language of thought hypothesis

One of the pre-philosophical characteristics of content described earlier was its systematicity. If one is capable of thinking that John loves Mary then one must also be capable of thinking, or understanding the thought, that Mary loves John. One may not actually think this—as love may be unrequited—but one must be able to think it. Similarly, if one can understand the thought that coal is black, one must be able to understand the thought that something else is black. One must be able to understand the possibility of the more general application of the concept of black. What explanation can be given of these systematic abilities?

The systematicity of thought and language

The most explicit answer to these questions is provided by the language of thought hypothesis. According to this, mental

representations are symbols in a language of thought with both a compositional syntax and semantics. The most direct motivation for this thesis is a swift response to the explanation of the systematicity of mental content. Such systematicity is also apparent in language. A language capable of representing the fact that John loves Mary (in English, by the sentence 'John loves Mary') is also capable of representing the fact that Mary loves John, by reversing the symbols for John and Mary in the representing sentence. So one obvious explanation for the systematicity of thought is to say that thinking comprises the manipulation of inner representational symbols in a language of thought. Thus as well as granting that the thought that John loves Mary is a structured thought, in accordance with Evans' Generality Constraint, the language of thought hypothesis claims that the internal bearer of this content is a structured or compositional symbol in an inner language.

Thus complex mental representations can be built up from more basic ones in a way that resembles the construction of complex from simple ideas in Locke's account of the mind.

The vehicles of content are structured

Fodor makes the distinction between the language of thought (LOT) hypothesis and other theories such as functionalism in the following passage:

Practically everybody thinks that the *objects* of intentional states are in some way complex: for example, that what you believe when you believe that ... P & Q is ... something composite, whose elements are—as it might be—the proposition that P and the proposition that Q.

But the (putative) complexity of the *intentional object* of a mental state does not, of course, entail the complexity of the mental state itself. It's here that LOT ventures beyond mere Intentional Realism ... LOT claims that *mental states*—and not just their propositional objects—*typically have constituent structure*. *Psychosemantics* (p. 136)

Such a claim ensures that those able to think the first thought are able to think the second thought because of the systematic relations between compositionally structured mental representations. So much then for the explanation of the systematicity of thought. But what of the rational structuring of thought? The systematicity of mental representations explains the fact that if one is able to think 'P and Q' one must also be able to think 'P' but not that one *ought* to think it. The language of thought hypothesis, however, also has an explanation of the rational or normative nature of our thinking.

The computer metaphor again

The computer provides an influential metaphor. Computers are machines that can be programmed to manipulate symbols so as to respect the semantic or rational relations of the contents that they encode. This manipulation, encoded in the lowest levels of computer programming languages, depends on causally relevant properties of the symbols, which, by design, correspond to their

syntax. Thus the harmony of the causal properties of symbols and the rational properties of contents is effected by the syntax of the symbols that encode those contents. This suggests that language of thought versions of representationalism can make use of a similar account to explain the rational structure of thought.

On this account, mental representations comprise symbols in a language of thought whose syntax governs their causal interactions such that the demands of rationality are respected. In other words, the mind is a computer whose computations involve the manipulation of internal symbols. Thinking comprises the causal manipulation of internal symbols according to a system of laws that matches the rational structure of the contents of thoughts encoded by the symbols.

Representational Theory of Mind and functionalism

The relation between Fodor's RTM, based on a language of thought differs from functionalism. Fodor thinks that the difference between hoping something and fearing it can be explained as a difference in the functional role of a mental representation. But he does not think that content of the attitude—what is hoped or feared—can be explained or determined by functionalism. That is instead explained as resulting from the combination of tokens in a language of thought.

Inner language or inner maps and models?

While the language of thought hypothesis is the most explicit representationalist attempt to explain the systematicity of thought, not all forms of representationalism are committed to it. An alternative explanation relies not on an inner linguistic structure but a structure based upon inner maps or models. Briefly, it has been argued that maps can also explain the systematicity and rational structure of thoughts. The same mapping conventions and symbols that enable a representation of Cambridge as to the east of Oxford could also represent Oxford as to the west of Cambridge. There is a systematicity in such mapping relations. An explanation of the rational structuring of thoughts based on maps is less clear than the language of thought explanation. But there is no reason in principle why a similar kind of story might not be developed. The harmony of the causal properties of map symbols and the rational properties of contents could be effected by a kind of spatial syntax of the mapping symbols, which, on this account, encode those contents. In what follows we will ignore the differences between language of thought and mental model theories of the mind.

There's still something missing!

Both the language of thought and the mental models hypotheses attempt to explain the structure of mental content through the structuring of internal states. But while these attempt to explain the 'syntactic' properties of thought (literally the syntax in the case of the language of thought and an 'internal' structure akin to syntax in the case of mental models), they do not explain the intentional or semantic properties of thought. That is, they do

not explain how mental representations—thought of as second order states of the brain—can be *about* anything. The next two sessions will examine two different theoretical additions that attempt to explain how inner representations can be about the world. The first is a straight causal theory of reference. The second (in Session 5) is a teleological theory.

Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What are some of the strategic options for explaining intentionality?
2. What are the relative priorities of mental and linguistic phenomena and what approach is generally favoured by reductionist accounts.
3. What in outline is Fodor's RTM and how does it encode propositional attitudes and how does it relate to a functionalist account of mind?

Session 4 Descriptive causal accounts of content

Semantics and the causal theory of reference

The representationalist picture of mind outlined so far postulates ontologically independent internal mental states standing in systematic causal relations one to another. They are 'ontologically independent' because their existence is not tied to any states of affairs in the outside world. (What is an ontologically *dependent* state? Think of the marriage of Posh and Becks. That state—their marriage—turns for its existence on the existence at some time of both parties. It is an essentially relational state.) But the very fact of their independence of the world raises a key problem: Why think of mental representations as representations of anything? Why think that they possess intentionality or semantic properties?

This question can be put in the context of the flow diagrams of the first reading in the chapter. At the heart of the modular and functional account of speech recognition and understanding was a 'semantic system'. On the assumption that such a system is thought of as a symbolic processor, operating on mental representations in the broad way Fodor describes, how does anything that goes on in there count as being about anything in the outer world? How could meanings be stored up in such a system as well as representations of words or the supposed *vehicles* of content? To answer these questions Fodor combines his account of RTM so far described with a causal theory of reference.

Fodor's RTM adds to the elements so far outlined a further claim. In addition to standing in causal relations to one another, mental representations also stand in causal relations to items in the external world. It is in virtue of their external causal relations that mental representations have intentional content.

Two roles for causal relations

In his book *The Representational Theory of Mind* the Australian philosopher K. Sterelny (1990) summarizes this idea as follows:

'Thoughts are inner representations; thinking is the processing of inner, mental representations. These representations have a double aspect... [T]heir role within the mind depends on their individualist, perhaps their syntactic, properties... [And] they are representations in virtue of relations with the world. Mental states represent in virtue of causal relations of some kind with what they represent.' (p. 39)

The basic idea is that mental representations gain their content by standing in suitable causal relations to items in the outside world. But simply adding additional external causal relations to the internal causal relations between mental representations does not explain how they come to possess intentionality.

Normativity and the disjunction problem

Representationalists do not generally attempt to tackle this problem directly. The problem that is generally acknowledged by representationalists is the problem of the *normativity* of beliefs. The assumption is something like this. If a representationalist theory of content cannot account for false beliefs, there is something fundamentally wrong with it. If the theory is not so falsified, then that may be some indication that it is indeed a theory of content. The problem of normativity is this. Beliefs can be both correct and incorrect. But a crude causal theory threatens to make false beliefs impossible if it simply identifies the content of a belief with what actually causes it. The theory will instead make all contents disjunctions of their causes. Thus the problem is often called the 'disjunction problem'.

To take an example. Suppose we want to be able to explain how it is possible to have the false thought that there is a cow in front of one when in fact there is rather a plump horse. (Any theory that cannot allow such false thoughts is a bad account of fallible human thinking.) Suppose also that we adopt a representationalist assumption that such thoughts are in fact internal states of people (states of their brains perhaps). Then we need an account of how such and such a brain state *means* cow even though it is caused in this case by a plump horse. So a simple theory that says that a brain state means whatever causes it cannot account for this false thought.

There are two general approaches that are designed to meet this problem. Both work by adding an extra ingredient to the simple causal link. The first is to modify a pure descriptive theory. The second is to deploy the resources of biological teleological explanations. In the reading below Fodor briefly summarizes and criticizes one of each of these before putting forward his own

descriptive causal theory. (Teleological theories will be the subject of the next session.)

Dretske's pure descriptive theory

In his influential 1981 book *Knowledge and the Flow of Information*, Fred Dretske (1932–) an engineer turned philosopher proposed a solution to the disjunction problem. He argued that a *learning process* establishes the correct causal connection between mental representations and what they are about. After an initial learning process, anything that causes a mental representation but that lies outside the correlation established during the learning time corresponds to an erroneous 'firing' of the mental representation rather than adding to the content it encodes.

But as Fodor argues, if a horse at night causes the mental representation that encodes 'cow'-thoughts to be tokened after the learning period, then presumably it would have caused it in the training period had it occurred then. Thus as one needs a counter-factual reading of the training period, this destroys the contrast between truth and error that this account was supposed to underpin.

Simple teleosemantics

Millikan's teleological reductionist account of mental content will be the subject of the next session. Fodor characterizes a slightly different form of that approach, which, for clarity, could be called a simple teleological solution. (See for example Papineau *Reality and Representation*, 1987, chapter 4.) Such a theory attempts to solve the disjunction problem by saying that the correlation that determines the content encoded by a mental representation is the causal connection that would occur under *optimal circumstances* where these are explicated through natural selection.

Fodor raises three objections to simple 'teleosemantics' (teleological semantics):

1. Why should what is optimal for belief fixing as far as natural selection is concerned be optimal for making them true? What about the selective advantage of fast but merely approximate thinking? And why should the conditions for optimality be the same for all beliefs? If not then the kind of circumstance in which mental representations are appropriately caused may have to be specified *after* settling their content.
2. Teleology merely replicates without solving the disjunction problem. Frogs could be described as responding to either flies or to little black things because in their environment the two coincide. Frogs of either type would survive equally well on earth. There is no material in actual historical natural selection to discriminate which is the *correct* intentional story.
3. This kind of account—which specifies conditions under which the causation corresponds to correct use—does not permit the *robustness* of meaning: that many things can cause me to think of cows quite correctly which are not cows (e.g. milk).

EXERCISE 5

(30 minutes)

Read the extract from:

Fodor, J.A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press, pp. 106–108

Link with Reading 24.4

- ◆ How satisfactory is Fodor's solution to the 'disjunction problem'?

Fodor's descriptive theory: asymmetric dependence

Fodor's preferred solution to the disjunction problem is to distinguish between causal connections that are constitutive of content and those that are not, in terms of asymmetric dependence. A type of mental representation has the content 'cow' if cows cause it to be tokened in the 'belief box' of a thinker and if those occasions on which it is caused by non-cows depend asymmetrically on the connection to cows. (Thus if the former connection had not existed—because the mental representation had a different content—then the latter would not have existed either.) Occasions when non-cows cause the 'cow' mental representation to be tokened can now be counted as errors.

Objections to Fodor's theory

Having now sketched out Fodor's positive theory we will turn, again rather briefly, to some standard objections. These can be divided between those that are more technical objections to the theory and for which some twiddling with the theory might be sufficient response and those that are fundamental objections and that aim to show that there is something inadequate about this whole approach to content, intentionality, or semantics.

Note first that a number of obvious objections to Fodor's theory can be avoided by making sure the dependence is specified in terms of *lawlike* connections between *properties* in the world. Thus a mental representation stands for cows if the property of cow-hood generally causes it. This allows, for example, for a mental representation to stand for unicorns even if it has never actually been tokened in the presence of a unicorn. It also helps avoid the following objection. If the mental representation that stands for horses has, as a matter of fact, only been produced by small horses—because these are the only horses so far experienced—does it really stand for small horses? Fodor replies that it does not providing it *would have been* caused by large horses had there been any.

Cats and robots

Lynne Rudder Baker raises an objection that introduces a disjunction back into the content encoded by mental representations based on an hypothetical example of cats and robot cats (Baker, 1991).

Sally has learnt the meaning of 'cat' from exposure to robot cats. (Thus her mental representations have been reliably caused by robot cats.) Suppose she now sees a normal biological cat. As it looks identical it will also cause a tokening of the 'cat' mental representation. What content does that representation encode then? Baker suggests the following options:

1. It correctly represents a cat and all the past tokens misrepresented robot cats as cats. But surely this gets the dependence relation the wrong way round?
2. It misrepresents this cat as a robot cat. But what of counterfactuals about what would have happened if she had seen this earlier?
3. It is disjunctive. Fodor opts for this. But this makes disjunction widespread. Most of us cannot distinguish horses and mules, for example. Also if it is disjunctive then Sally cannot later recognize her earlier application as an *error* when she finds out there are two sorts of thing. How do we get to asymmetric dependence from the earlier disjunction?

Causality and normativity

There are a number of similar arguments in the literature. Each tries to show that Fodor does not successfully solve the disjunction problem. But there is another more general problem highlighted by Godfrey-Smith (1989). He asks what resources a *purely causal* theory has for distinguishing between the independent causal relation that determines the content of a mental representation and those dependent causal relations that correspond to error?

Consider a mental representation that is caused by normal-looking horses, athletic cows, mucky zebras, and so forth. One obvious interpretation of this is that the representation encodes horse-thoughts and that the connections between it and some cows and zebras asymmetrically depends on the connection to horses. But there is another interpretation that is equally plausible given only the facts about causal connections. That is that the mental representation encodes a disjunctive content including normal looking horses, and some cows and zebras. There is, after all, a *reliable* causal connection between those animals and the representation. It is only *given* the content of the mental representation that one can determine which connections are fundamental and which are dependent, which would hold in nearby possible worlds and which would not. What this suggests is not just that there is a problem with Fodor's particular solution but that there is something generally wrong with attempts to reduce intentional notions to purely causal ones. Something is omitted by the causal account. It cannot determine which the *correct* application is.

Another way of looking at this last point is like this. Although Fodor may be correct to say that error depends on truth, the asymmetric dependence theory does not by itself explain how mental representations come to encode contents in the first place. If a mental representation has a certain meaning, then its

application in errors will depend logically on a prior veridical use. But the causal story is itself insufficient to explain the antecedent. How is such encoding brought about? Asymmetric causal relations may be necessary but not sufficient for content.

To repeat, none of these objections would be accepted as decisive by Fodor. Think of his approach as a research programme in the Lakatosian sense (see Chapter 16). If so it is characterized by a central core of assumptions about how intentionality is to be explained and by less central hypotheses that spell this out. The technical objections threaten the latter outer hypotheses and thus might be countered by technical fixes of one sort or another. The more fundamental objections challenge the central core. But whether there are seen to be decisive will turn in part on what other options are available.

In general, however, the tide in reductionist philosophy of thought has turned against pure descriptive theories in favour of teleological evolutionary theories. We will now turn to one of these to see if it could explain how what might be processed in a 'semantic system' could possess intentionality.

Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What needs to be added to the comparison with a computer for a reductionist account of intentionality?
2. What in broad outline is Fodor's solution and what further problem does he think it faces?

Session 5 Teleological causal accounts of content

One response to the charge that purely causal theories cannot account for the normativity of content is to think that a further ingredient has to be added. The most plausible explanation of what ensures that the causal connections between mental representations accord with their rational relations is the process of evolution. This is the motivation for a teleological causal theory. Although it adds an extra ingredient to account for the normativity of thought—the disjunction problem again—it is still part of a reductionist programme. It still seeks to explain in naturalistic terms derived from biological science how there can be meaning in the world, how there can be intentionality.

The most thorough version of a teleological account has been developed the American philosopher of thought and biology Ruth Garrett Millikan. The locus classicus is her book *Language, Thought and Other Biological Categories* (1984) and *White Queen Psychology* (1993) contains a number of more recent papers.

EXERCISE 6

(45 minutes)

Read the extract from:

Millikan, R. (1995). Biosemantics: explanation in biopsychology. In *Philosophy of Psychology* (ed. C. Macdonald and G. Macdonald). Oxford: Blackwell, pp. 253–276. (Extracts pp. 253, 255–258)

Link with Reading 24.5

This reading provides a concise introduction to Millikan's teleological explanation of content. Think what extra resources biological teleological explanations provide for 'naturalizing' content. What is the source of this conceptual extra?

Biological function

Teleological theories of content appear to have an important extra resource for explaining the normativity of content over and above those available to pure descriptive theories. They can employ the notion of a natural, proper, or biological function. There is some debate about how precisely to define such a function. Roughly speaking, however, it is that function that a particular trait of an organism exemplifies and explains the evolutionary success and survival value of that trait. Crucially, biological functions are distinct from dispositions. The biological function of a trait and its dispositions can diverge. Engineering limitations might cause the actual behavioural dispositions of a trait to diverge from the biological function it thus only partially exemplifies. The divergences might themselves be life threatening and play no positive part in explaining the value of the trait. The best explanation of the survival of that organism and those like it cites the *function* that helped propagation or predator evasion, for example, and not those aspects of its behavioural dispositions that diverged unhelpfully from it.

of and for

This point is sometimes put by saying that what matters is not what traits or dispositions are selected, but what *function* they are selected *for*. The distinction between 'selection of' and 'selection for' can be illustrated by the example of a child's toy (from Sober, 1984). A box allows objects of different shapes to be posted into it through differently shaped slots in the lid. The round slot thus allows the insertion of balls, for example. It may be that the actual balls allowed through or 'selected' in one case are all green. But they are *selected for* their round cross-section and not their green colour. Millikan stresses the fact that the biological function of a trait may be displayed in only a minority of actual cases. It is the function of sperm to fertilize an egg but the great majority of sperm fails in this regard (see Millikan, 1984, p. 34).

As biological functions can diverge from mere dispositions, teleological causal explanations of content have an extra resource to explain normativity. The distinction between correctness and incorrectness in the tokening of a mental representation can be

defined by reference to its functioning in accord with its biological function. Its function may be to be caused only by cows, for example, and not by plump horses. Thus if it is caused by a plump horse that still does not make it *mean* plump horse. It means 'cow', although it can also be caused by horses when the lighting is poor. This, supposedly, solves the disjunction problem with respect to the content of mental representations.

A reduction of rationality?

It also seems that an appeal to biological function can also explain how mental representations stand in rational relations. Indeed, Millikan elsewhere goes further by suggesting that, given a teleological account, logic will become the first *natural* science (Millikan, 1984 p. 11). That is, teleology will underpin a psychological explanation of logic and rationality. What makes one thought logically follow from another or one number the correct result of a calculation depends, ultimately, on abstract rules that humans apply to experience and that have had survival value.

Teleology and cognitive psychology

Millikan's teleological theory differs from Fodor's in that it adds in an extra ingredient for explaining the semantics of mental states: their proper or biological function. But it still broadly shares Fodor's goal of explaining how the semantics or content of mental states arises from natural processes that can be described in non-intentional terms. As such, it could be combined with the sort of computational or representationalist account favoured by cognitive psychology. The 'semantic system' could be described using evolutionary terms in such a way as to answer the question raised earlier about why one should think that anything that goes on in it is really semantic or world-involving. The states processed by such a system are semantic because that is their biological function.

Objections to teleosemantics

Unsurprisingly, Millikan's account faces criticism. We will focus on a criticism that concerns the reductionist aspirations of Millikan's theory.

Fast thinking

There are, however, two fundamental objections to a teleological approach. One has already been raised by Fodor (see above). It is that a teleological approach presupposes that the requirements of survival and rationality must go hand in hand. It is far from clear that this is actually so or that it must be so. In some circumstances it seems much more likely that quick approximations will contribute more to survival than a slow derivation of the correct answer. But even if, by chance, rationality, and survival coincided in the evolutionary history of animals on earth, this would be a contingent matter and thus undermine the necessity of logic.

A general tension in teleosemantics

The second flaw stems from the real source of the teleological theory's apparent ability to account for normativity. As already suggested, this results from the fact that biological functions are

not pure dispositions. The extra normative ingredient derives from this difference. Biological functions are defined by reference to what *explains* the selection of a trait. It is this idea of explanation that introduces the extra normative character of biological function. But two points are important here, which pull in different directions:

What matters in this explanation is what traits are selected *for* rather than just what the traits are that are actually selected. This follows from the fact that even traits that are favoured by evolutionary selection may have design flaws. Some behavioural dispositions that are the causal consequences of a successful trait might themselves have deleterious effects on survival.

But nothing should be counted part of the biological function of a trait which has not contributed to the causal explanation of survival of that trait. This follows from the fact that biological teleology is still really the result of normal non-teleological causal processes.

Peacocke on reduced content

This latter feature forms the basis of a criticism developed by Peacocke in *A Study of Concepts* (1992, pp. 129–132). Taking the case of a belief forming mechanism, he argues that the only consequences that can explain the success of such a mechanism are the consequences of beliefs formed that have a causal impact on organisms that have it. But the truth of the causally relevant consequences of a belief can fall short of the truth of a belief itself. Thus:

In a nutshell, the problem of reduced content is this: how is the teleological theorist to block an incorrect assignment of content to beliefs, namely one that requires for its truth merely the truth of all the logical consequences of *p* that have a causal impact on the thinker, rather than the stronger condition of the truth of *p* itself?
Peacocke (1992, p. 130)

Millikan's reply

Millikan's response is to stress again the first of the points above: what matters is not *what* mechanism is selected but what it is selected *for* (Millikan, 1995). She argues that the best explanation of the presence of any such belief forming mechanism will not ascribe to it a function of the limited form that Peacocke suggests. Its function is to represent *p* and not merely to select the consequences of *p* that have causal impact. But this response points, however, to a more fundamental problem. It turns on an argument from Wittgenstein mentioned in chapter 14 and discussed more fully in chapter 25 but the key point can be sketched out now.

Millikan and Wittgenstein

A teleological account of function is a form of *interpretational* theory because the characterization of the function that explains the survival of a trait is in effect an interpretation of the past behaviour. Past behaviour comprises signs to be interpreted. But

as will be made clearer in the next chapter, there is no limit on the different way a set of signs or one's own past behaviour can be interpreted. Like the interpretation of signs, such behaviour is *consistent* with an unlimited number of possible functions or rules. What ensures the determinacy of biological function—what selects just one of the rules—is an explanation of the presence of a trait couched in intentional terms that interprets what the trait is for. But finite past behaviour can be explained as exemplifying many different or 'bent' functions or rules, all of which would have been equally successful in the past. (The idea of a bent rule will be re-explained in the next chapter.)

Millikan dismisses these possible alternatives by stressing, this time, the second of the points above. The explanation of a trait turns on what caused it to survive in the past. Taking the case of the rules that govern a hoverfly's mating behaviour she argues:

[The 'bent' rule] is not a rule the hoverfly has a biological purpose to follow. For it is not because their behaviour coincided with *that* rule that the hoverfly's ancestors managed to catch females, and hence to proliferate. In saying that, I don't have any particular theory of the nature of explanation up my sleeve. But surely, on any reasonable account, a complexity that can simply be dropped from the explanans without affecting the tightness of the relation of explanans to explanandum is not a *functioning* part of the explanation.
Millikan (1993, p. 221)

The claim is that 'bent' rules introduce additional and unnecessary complexities that can be dispensed with without damaging the explanation of the success of a biological trait. But the claim that bent rules are overcomplex and can be rejected either presupposes a particular interpretation of past behaviour for comparison—in which case it is question-begging—or that judgements can be made about the simplicity of rules from an objective Platonist perspective. And that is an idea that Wittgenstein has demolished (we saw some of this in Chapter 14).

A dilemma for natural selective explanation

It is worth noting that in the case of the mental representations of observable states of affairs, the suggestion that their biological function can be specified objectively via the explanation of their survival is more attractive. But this is misleading because such explanation helps itself to natural kinds in the explanation of what it is that mental representations track. In the context of explaining content, such explanation faces a dilemma:

- ◆ *Either*: natural kinds are identified with the extension of mental representations, in which case they cannot be invoked to explain the content of mental representations.
- ◆ *Or*: they are identified with classification-independent groupings or patterns in the world. The idea here is that thought tracks groupings that already exist independently of human judgement. This resembles the Platonist theory that Wittgenstein criticizes in that it locates the standard of correctness for such judgements outside human practice. Thus it cannot account

for the normativity of judgements about what does and does not belong to the classification. But a teleological theory faces a more specific difficulty. There is no reason to believe that the biological function of a representation will be to track such natural kinds as opposed to functional kinds. The categories that are useful for an organism may not align with natural kinds.

Natural and functional kinds

An example that has been used elsewhere helps emphasize this second point. Consider a mental representation that encodes the content poison. This seems to be a functional concept in that it denotes any substance that is toxic to the organism in question. But it is possible that in the circumstances in which it had survival value, all the causes of that representation belonged to the same natural kind, the same biological species. The problem now is what, according to a teleological theory, distinguishes representations that encode what are, pre-philosophically, functional kind contents from natural kind contents? It seems on reflection that even in the case of mental representations that concern detectable states of affairs, the teleological theory is no improvement on the pure descriptive causal theory. Neither can account for the normativity of content.

A deeper objection to reductionism

There is a further general Wittgensteinian objection to both descriptive and teleological causal theories of content. Both attempt to add something to internal mental states in order to explain their meaning or life. But what they add does not make the right sort of difference. The asymmetric dependence of the tokening of internal states on types of states of affairs does not explain why those states should be thought of as representational. Equally, the fact that internal states—which have the fortuitous causal consequences for movement—lead to natural selective advantage, does not explain why they are representational.

Chapter 23 noted that functionalism alone did not succeed in explaining the intentionality of mental states. Functionalism characterized them as merely a system of causal ‘pushes and pulls in the head’. These ‘horizontal’ causal relationships could be sufficient to explain behaviour without the need to think of additional ‘vertical’ relationships of representation. Why think of the internal states so characterized by functionalism as about anything?

The teleological theory is supposed to provide an answer to precisely this question. The suggestion is that once an account has been given of ‘vertical’ causal relations between internal states and states in the world, the problem of naturalizing content will have been solved. But it does not. All it provides is a further specification of which internal causal states have causally advantageous consequences. It does not explain why the ‘vertical’ causal relations that it provides should, in addition to fulfilling this causal role, fulfil an intentional role.

The consequences for a cognitive psychological explanation of semantics

This chapter has examined philosophical explanations of mental content that have adopted one of two forms of reductionism. Neither has been successful at reducing intentional notions to non-intentional causal relations between internal mental representations and parts of the world. This, however, is the assumption behind cognitivist explanations of mental content. The idea is that the computer can be used as a metaphor for the mind in order to explain cognitive abilities in information-processing terms. But this idea faces a severe objection because computer-based information processing requires underlying states that can be processed. Real computers are so designed that their internal states possess causal properties that are designed to match syntactic or programming properties. But once one thinks of human mental states as explained by free-standing internal mental representations, this forces the question: Why think of these as having semantic properties? Simply postulating a ‘semantic system’ at the heart of mental processing does nothing to answer this question.

Both Fodor and Millikan attempt to resolve this issue by describing links between internal states and the external world, either through a purely causal connection or a teleological causal connection. Both of these attempts to explain intentional connections through causal connections, however, face grave difficulties. And this casts doubt on the very idea of explaining understanding through an internal ‘semantic system’. But if this approach cannot work, what alternative is there?

The underlying motivation for reductionism identified earlier in this chapter was that without it, content could not form an unmagical part of the natural world. In the next chapter we will examine an alternative approach that takes the failure to reduce content as in no sense undermining a properly naturalistic sense of intentionality. But on the other hand, trying to explain semantic properties at a lower level using subpersonal systems has no prospect of success. Only whole people can have semantic states.

Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What other broad approach to the reductionist explanation of intentionality is there aside from Fodor’s causal approach?
2. What role does biological function have in such accounts? What is the importance of the distinction between selection of and selection for?
3. What problems does it face?
4. What consequences do these difficulties raise for cognitive psychiatric theories of meaning?

Reading guide

- ◆ Useful introductions to the philosophy of content or intentionality can be found in the general philosophy of mind textbook by Braddon-Mitchell and Jackson (1996) *Philosophy of Mind and Cognition* (chapters 10–11), as well as in textbooks on the philosophy of content specifically. See, for example, Luntley (1999) *Contemporary Philosophy of Thought*.
- ◆ For an introduction to the philosophy of language sympathetic to a reductionist, representationalist approach see Devitt and Sterelny (1999) *Language and Reality: an introduction to the philosophy of language*.
- ◆ Criticism of the very idea of a meaning-driven folk psychology in favour of eliminativism is found in P(atricia) Churchland's (1986) *Neurophilosophy*, and Stich's (1983) *From Folk Psychology to Cognitive Science* and Stich's (1992). What is a theory of mental representation? *Mind* 101: 243–261.

Representationalism

- ◆ Representationalism is set out in Sterelny's (1990) *The Representational Theory of Mind*.
- ◆ The language of thought is discussed in the appendix to Fodor's (1987) *Psychosemantics*.
- ◆ For a Wittgensteinian critique of representationalism see Thornton's (1998) *Wittgenstein on Language and Thought*.
- ◆ And for representationalist approaches in psychology see Thornton (2002).
- ◆ Ruth Millikan's work is set out fully in *Language Thought and other Biological Categories* (1984) and *White Queen Psychology* (1993). Concise introductions are G. Macdonald 'The biological turn' and Millikan 'Biosemantics: explanation in biopsychology', in C. Macdonald and G. Macdonald *Philosophy of Psychology*, pp. 238–276.
- ◆ For a more accessible introduction to a biological teleological approach to meaning ('teleosemantics') see Papineau (1987) *Reality and Representation*, and (1993) *Philosophical Naturalism*.

Clinical and interdisciplinary work

- ◆ Clinical work in a cognitive neuropsychological framework can be found in Caramazza (ed.) (1990) *Cognitive Neuropsychology and Neurolinguistics*.
- ◆ A philosophical attempt to set out the basis of a reconciliation of causal and information rich explanation for mental health practice is described in Bolton and Hill (1996) *Mind Meaning and Mental Disorder*.
- ◆ Interesting interdisciplinary work can be found in Carruthers and Boucher (1998) *Language and Thought: interdisciplinary*

themes, and in Bechtel and Graham (ed.) (1999) *A Companion to Cognitive Science*. Both of these contain discussions of modularity in cognitivist accounts of the mind.

References

- Baker, L.R. (1991). Has content been naturalised? In *Meaning in Mind: Fodor and his critics* (ed. B. Loewer and G. Rey). Oxford: Blackwell, pp. 17–32.
- Bechtel, W. and Graham, G. (ed.) (1999). *A Companion to Cognitive Science*. Oxford: Blackwell.
- Bolton, D. and Hill, J. ([1996] 2nd edn 2004). *Mind Meaning and Mental Disorder*. Oxford: Oxford University Press.
- Braddon-Mitchell, D. and Jackson, F. (1996). *Philosophy of Mind and Cognition*. Oxford: Blackwell.
- Brentano, F. ([1874] 1995). *Psychology from an Empirical Standpoint*. (trans. Antos C. Rancurello, D.B. Terrell, and Linda L. McAlister). London: Routledge.
- Caramazza, A. (ed.) (1990). *Cognitive Neuropsychology and Neurolinguistics*. Baltimore, MD: Johns Hopkins University.
- Carruthers, P. and Boucher, J. (1998). *Language and Thought: interdisciplinary themes*. Cambridge: Cambridge University Press.
- Churchland, P.S. (1989). *Neurophilosophy*. Cambridge, MA: MIT Press.
- Devitt, M. and Sterelny, K. (1999). *Language and Reality: an introduction to the philosophy of language*. Oxford: Blackwell.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press.
- Ellis, A. (1996). *Human Cognitive Neuropsychology: a textbook with readings*. Hove, UK: Psychology Press.
- Evans, G. (1982). *The Varieties of Reference*. Oxford: Clarendon.
- Fodor, J.A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
- Fodor, J.A. (1991). Propositional attitudes. In *The Nature of Mind* (ed. D. Rosenthal). Oxford: Oxford University Press, pp. 325–338.
- Godfrey-Smith, (1989). Misinformation. *Canadian Journal of Philosophy*, 19: 533–550.
- Grice, H.P. (1969). Utterer's meaning and intentions. *Philosophical Review*, 78: 147–177.
- Luntley, M.O. (1999). *Contemporary Philosophy of Thought*. Oxford: Blackwell.
- Macdonald, G. (1995). The biological turn. In *Philosophy of Psychology* (ed. C. Macdonald and G. Macdonald). Oxford: Blackwell, pp. 238–251.

OUP Copyright

- McGinn, C. (1989). *Mental Content*. Oxford: Basil Blackwell.
- Millikan, R. (1984). *Language Thought and other Biological Categories*. Cambridge, MA: MIT Press.
- Millikan, R. (1993). *White Queen Psychology*. Cambridge, MA: MIT Press.
- Millikan, R. (1995). Biosemantics: explanation in biopsychology. In *Philosophy of Psychology* (ed. C. Macdonald and G. Macdonald). Oxford: Blackwell, pp. 238–276.
- Millikan, R.G. (1984). *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.
- Millikan, R.G. (1993). *White Queen Psychology*. Cambridge, MA: MIT Press.
- Papineau, D. (1987). *Reality and Representation*. Oxford: Blackwell.
- Papineau, D. (1993). *Philosophical Naturalism*. Oxford: Blackwell.
- Parkin, A.J. (1996). *Explorations in Cognitive Neuropsychology*. Oxford: Blackwell.
- Peacocke, C. (1992). *A Study of Concepts*. Cambridge, Mass.: MIT Press.
- Sobere (1984). *The Nature of Selection*. Chicago: University of Chicago Press.
- Sterelny, K. (1990). *The Representational Theory of Mind*. Oxford: Blackwell.
- Stich, S. (1983). *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
- Thornton, T. (1998). *Wittgenstein on Language and Thought*. EUP.
- Thornton, T. (2002). Thought insertion, cognitivism and inner space. *Cognitive Neuropsychiatry*, 7: 237–249.

OUP Copyright