

Meaning

A SLIM GUIDE TO SEMANTICS

Paul Elbourne

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© Paul Elbourne 2011

The moral rights of the author have been asserted
Database right Oxford University Press (maker)

First published 2011

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloguing in Publication Data
Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
Clays Ltd, St Ives plc

ISBN 978-0-19-958583-0 (Hbk)
978-0-19-969662-8 (Pbk)

1 3 5 7 9 10 8 6 4 2

To Siofra

Preface

The ability to communicate linguistically with each other in the extraordinarily rich way that we do is a distinguishing feature of the human species. To communicate linguistically is to convey meaning. This book offers an introduction to semantics, the discipline that analyses meaning. Semantics asks questions like

'What is meaning?'

'How do meanings of words combine with each other to give us meanings of sentences?'

and

'Do the meanings of the words in our languages influence what thoughts we can think?' All of these questions will be addressed in this book.

To be precise, this book is about natural language semantics, which is the analysis of the meanings of words and sentences in natural languages like English and Japanese. It will have nothing to say about the semantics of computer programming languages and other artificial languages, important though that topic is. And it will regrettably have nothing to say about the meaning of life, important though that topic arguably is too. Natural language semantics is a peculiar discipline in that it is carried out under the collective aegis of three larger subjects: linguistics, psychology, and philosophy. This book looks at theories from all three. Semantics is also notable for the amount of controversy involved, on everything from small details to the most basic foundations of the field. Unlike some other introductory texts, this book will not shy away from exploring disagreements and difficulties.

I am very grateful to John Davey, of OUP, for his encouragement and for all kinds of advice; to Lera Boroditsky for confirming some details of her experiments; and to Joanne Dixon, Lee Jackson, and Siofra Pierse for reading various drafts of the manuscript and telling me what was incomprehensible and what was not.

1

Definitions

Words are traditionally supposed to have meanings. Indeed it is widely supposed to be possible to *define* words' meanings. Whole books, called dictionaries, are devoted to listing the definitions of words; and philosophers from Socrates (469–399 BC) and Plato (429–347 BC) onwards have devoted obsessive attention to pinning down the meaning of philosophically interesting words like *knowledge*, *truth*, *justice*, and, indeed, *meaning*. It is important for anyone embarking upon the study of semantics to realize, however, that defining the meaning of a word is an enterprise of almost inconceivable complexity. Despite 2,400 years or so of trying, it is unclear that anyone has ever come up with an adequate definition of any word whatsoever, even the simplest. Certainly the definitions in dictionaries are the merest hints, and are sometimes flat out wrong.

Before we look at some examples of attempted definitions, it will be useful to formulate a standard by which we might appropriately judge them. Suppose I define *chair* as 'item of furniture'. It is clear, I think, that my definition is faulty. Why? Because there are plenty of things that are items of furniture that are not chairs—tables, desks, footstools, and so on. My definition is too lax, in the sense that it includes too many things. Suppose, on the other hand, that I define *chair* as 'throne'. My definition is once again flawed. All thrones are plausibly chairs, but there are lots of chairs that are not thrones. My definition is now too strict, in the sense that it excludes too many things. A good definition of the word *chair*, it seems, must be neither too strict nor too lax; in other words, it must pick out all and only the things that are chairs. And similarly for definitions of other words.

How well do dictionary definitions of *chair* measure up on this score? Let us look at a few and find out. The *Collins Pocket English Dictionary*, one of the more respected and well known smaller dictionaries of English, in its 2008 edition, defines *chair* as 'a seat with a back and four legs, for one person to sit on.' Does this pick out all and only the things that are chairs? Why, no, it does not. If that is not immediately obvious to you, think about the chairs in which the office workers of today can be found sitting at their desks. Some people do of course use a seat with a back and four legs for this purpose. But many are to be found swivelling around in a seat that rests on one central column that splays out near ground level into five or six separate castor-bearing feet. However you do the count, you cannot plausibly impute four legs to these devices; and yet they are indubitably chairs. So this definition is too strict, in the sense that it unjustly denies chairhood to many things that merit it.

Interestingly, the definition also seems to be too lax. Think back to the Diogenes Club of the Sherlock Holmes stories, a club in which no member is allowed to take the slightest bit of notice of any other member. Imagine that it has a garden adorned with ordinary garden benches. Two or three people could easily fit on each bench. But the club rules, we can well imagine, forbid any person to sit on a bench that is already occupied by another person. These garden benches, then, are seats with a back and four legs, for one person to sit on. But they are surely not chairs. (If it is relevant, we can imagine that the designers and manufacturers of the benches knew the use to which they would be put, so that no-one ever intended that these benches would be occupied by more than one person at any time.) So the definition of *chair* in the *Collins Pocket English Dictionary* is simultaneously too strict and too lax.

Perhaps you are thinking that it is unfair to pick on a 'pocket dictionary'. Such dictionaries, if they are to have any chance of actually fitting into people's pockets, will not have the space to include all the details about leg-count and occupancy that a larger dictionary might. So let us go to the opposite extreme. The *Oxford English Dictionary*, in its second edition of 1989, comprises twenty volumes and

21,728 pages, and takes up several feet of shelf-space. Let us see what it has to say on the subject of chairs:

A seat for one person (always implying more or less of comfort and ease); now the common name for the movable four-legged seat with a rest for the back, which constitutes, in many forms of rudeness or elegance, an ordinary article of household furniture, and is also used in gardens or wherever it is usual to sit.

The format of this entry is slightly complex, in that it seems to offer two alternative definitions: the phrases that come respectively before and after the semi-colon. The 'now' is perhaps to be taken as implying some historical development of the meaning of the word. So let us concentrate on the second, more up-to-date, definition, 'the movable four-legged seat with a rest for the back, which constitutes, in many forms of rudeness or elegance, an ordinary article of household furniture, and is also used in gardens or wherever it is usual to sit'. Does this succeed in picking out all and only the chairs? No. For all its elegance of phrasing and luxuriance of detail, it makes the same mistake about four-leggedness that the *Collins Pocket English Dictionary* made.

What if we emended these definitions to allow for different numbers of legs? It is not immediately clear how we should do so. When you think about it, it becomes obvious that chairs could come with all kinds of different numbers of legs. An avant-garde designer could easily promote a three-legged chair, a five-legged chair or a 100-legged chair. (For the latter case, imagine very thin legs, perhaps arranged in a ten-by-ten grid.) And think back to the swivelling office chair that I just described. Is it accurate to describe this kind of chair as having legs at all? Not for my money. And if you need any further convincing, imagine a solid cube of wood that reaches the customary height of a chair seat when placed on the ground, and imagine that it has a back like a chair. Such an object could, in fact, be a chair; but it would definitely not have legs.

So maybe we should charitably pass over the claims about legs. What about the other components of the *OED* definition? To start at the start, we have already seen difficulties with the claim that a chair

is a seat for one person: it introduces an inappropriate degree of laxity into the definition since it allows the garden benches at the Diogenes Club to be chairs. We have not yet observed that this condition is also too strict. I am sure that you have in fact seen perfectly ordinary chairs with two people on them, one person sitting on the lap of the other. Does this not raise a difficulty for the idea that chairs are seats for one person? Well, we could perhaps defend the dictionary definitions on this point by offering some kind of explication of the word *for*. Perhaps *for* here means something like 'designed for'. In other words, we could claim that, even if some people do utilize chairs in the peculiarly unnecessary manner just described, what the dictionary authors are driving at here is something like the intent of the designer of the chairs: chairs would be designed for one person, even if some irresponsible types do not respect this. But imagine a society in which, whether due to decadence or thrift, all chairs are designed to support two people (perhaps they are slightly reinforced) and are in fact generally used like this. Would these items of furniture still be chairs? Of course. The very fact that I can write without contradiction about chairs that are designed to support two people shows this. There is no internal contradiction in the idea of a chair that is designed to support two people. Compare the idea of a unicycle with two wheels, which does seem genuinely self-contradictory. This is evidence that the idea of being associated with *one* of something or other does form part of the meaning of the word *unicycle*, whereas it does not form part of the meaning of the word *chair*.

The technique that I just used to argue against chairs being seats for one, frivolous though it may seem, actually bears closer examination, for there is a moral hidden here. We must not confuse the meaning of a word with details about how the things that that word designates are in fact produced or used. We can tell this because we can use our common, everyday words to describe components of the most outlandish counterfactual situations. Semanticists distinguish between the *extension* and the *intension* of a term like *chair*: roughly speaking, the extension of *chair* is the set of all actual chairs, while the intension is the set of possible chairs, allowing for all the possibilities of bizarre science-fiction scenarios. The meaning of *chair*, whatever

it is, allows us to talk not only about actual chairs, but also about merely possible chairs.

The *OED* definition says that chairs are movable, which strikes me as a good generalization, but certainly not exceptionless, and almost certainly not part of the meaning of the word *chair*. Think of a daring architect who proposes a kitchen or a dining room in which the chairs are sculpted from the stone that forms the floor. There is nothing self-contradictory about the notion. If we abstract away from the *OED*'s talk of comfort, ease, rudeness, elegance and gardens, which forms part of the baroque splendour of the entry but does not increase our understanding of chairs, there are two ideas remaining: that of a seat, and that of a rest for the back. What do the *OED* authors mean by *seat*? Fortunately, we can look up the word in the *OED* and see. The most relevant sense in the article on *seat* seems to be 'Something adapted or used for sitting upon'. Now it might seem uncontroversial that chairs are things adapted or used for sitting upon. But even this is dubious, at least if the definition is taken to claim that all chairs have this property. Imagine a grand stately home open to the public. Much of the original furniture is still there, but a few chairs are missing. The owners decide to commission the construction of some chairs, replicas of the old ones, to fill in a couple of gaps around a great dining table. They do not intend that people should sit on them, however; in fact the whole ensemble is to be shut off behind a velvet rope and no-one is to be allowed to touch the chairs in question. In case it matters, we can further suppose that the designers and manufacturers know this. These chairs, then, are not adapted or used for sitting upon. And yet they are indubitably chairs.

This leaves the idea of a 'rest for the back'. If this is taken to imply that actual human backs have to touch chairs, it seems to be falsified by the scenario about the replica chairs in the manor house; and similarly if it is understood in a slightly weaker form, so as to imply only that the manufacturers or designers must intend this. It is not obvious that this phrase can be taken in any other way.

Overall, we have found that chairs do not have to have any of the properties ascribed to them by the definition of *chair* in the *Oxford*

English Dictionary. If we judge the definition as an attempt to pick out all and only the possible chairs, it fails. Other dictionaries perform no better.

Perhaps even the *OED* does not have enough space to go into all the necessary details that a good definition would require. Or perhaps its editors have not had enough time to find out what *chair* means: work on the dictionary only began in 1879, after all, which is quite recent in terms of the history of scholarship. What about philosophy? Philosophers, as I mentioned earlier, have devoted about 2,400 years to formulating the definitions of philosophically interesting words, and they can, and frequently do, devote whole books to just one such word. Have philosophers succeeded in defining a word after all this time? Not obviously. There may possibly be an accurate definition of a word lurking in some philosophical manuscript somewhere, but it is difficult to know what it might be, because there is no consensus among philosophers on any such case. On all the examples I mentioned earlier (*knowledge, truth, justice, meaning*), and many more, there is still controversy.

To get the flavour of the enterprise, let us consider the definition of *knowledge*. (Unfortunately, the word *chair* has come in for only limited philosophical analysis.) And to avoid getting tangled up in ambiguity (of which more later), let us concentrate on what is called *propositional knowledge*: knowledge that something is the case (for example, that snow is white), as opposed to knowing (or being acquainted with) a place or a person. For quite some time, it was thought that *knowledge* could be defined as 'justified true belief'. The first analysis of this kind, in fact, goes back to Plato's *Meno* (fourth century BC). Why should one think this? Well, it seems intuitively plausible that for you to know that snow is white you must at least believe it. Knowing is a kind of believing, perhaps with other conditions thrown in. Furthermore, if you know some proposition then that proposition has to be true. You cannot know that trepanning cures people of demonic possession, because it is not true that trepanning cures people of demonic possession. (Those in thrall to the intellectual charlatantry known as 'postmodernism' might seek to convince you that nothing is true. In most cases, however, the

question 'So is it *true* that nothing is true?' is enough to discombobulate them.) Why not stop here and say that knowledge is true belief? The reason is that it is possible to acquire true beliefs by accident, as it were, but we feel queasy about designating such beliefs as knowledge: a madman amidst his ravings might sincerely shout out some substantive and interesting true propositions that would be no better grounded than his belief that he is Napoleon. The requirement that the belief in question be justified somehow is meant to rule out this kind of thing from qualifying as knowledge.

So matters might have rested (I am simplifying the history somewhat) had it not been for the sublimely concise Edmund Gettier. Gettier is one of the most eminent living philosophers; but, rather splendidly, his entire published *oeuvre* consists of one three-page paper, an article from 1963 called 'Is justified true belief knowledge?' The answer to the question is no. Suppose, says Gettier, that Smith and Jones have applied for the same job. Before the result of their applications is announced, two things happen: Smith counts the coins in Jones's pocket and finds that they number ten; and the president of the company assures Smith that Jones will get the job. (We are not supposed to wonder why these things happen. This is a philosophical example, not a psychological novel.) Smith thus justifiably believes that Jones is the man who will get the job and that Jones has ten coins in his pocket. Being an impeccable logician he deduces that the man who will get the job has ten coins in his pocket. Now he is surely justified in believing this latter proposition: it is a watertight deduction from two things that he is already fully justified in believing. As it happens, however, Smith, not Jones, gets the job. And unbeknownst to himself, Smith too had ten coins in his pocket at the time that he counted Jones's coins and formed his beliefs. So it turns out that his belief that the man who would get the job had ten coins in his pocket was true. And it was also justified. But it is discomfiting in the extreme to say that Smith knew that the man who would get the job had ten coins in his pocket. So not all justified true belief is knowledge. A large part of the history of epistemology since 1963 has consisted of efforts to solve the 'Gettier problem', sometimes involving attempts to add some elusive fourth

property to the 'justified true belief' definition of *knowledge* and sometimes veering off in other directions; but there is no consensus in the field.

Perhaps you are now expecting me to come up with dazzling definitions of *chair* and *knowledge* that remedy the above deficiencies. I am afraid I must disappoint. The aim of this exercise has been to impress upon you the extraordinary difficulty of giving adequate definitions of words, even apparently humble ones. To give you a further taste of the difficulties that arise in this kind of exercise, let us examine some surprising facts about word meaning that have been pointed out by Noam Chomsky, the founder of generative linguistics and one of the leading figures in the 'cognitive revolution' of the 1950s and 1960s, which saw the foundation of modern cognitive psychology and artificial intelligence. In the discussion of *chair*, you may have remarked upon the important role that human intentions play in defining what seems at first to be a word for a straightforward physical object. Chomsky's observation is that this phenomenon is much more widespread than you might have thought, even in the case of words that do not denote human artefacts. If tea leaves have been deposited in your local reservoir by the proper authorities as a new kind of water purifier, what comes out of your tap will still be called *water*, even if (on one way of looking at it) it is an extremely mild tea; but if someone likes their tea very mild and dips a tea bag for just a split second into a cup of pure H₂O, the resulting liquid is tea and not water, even if it is chemically identical to the stuff that comes out of the tap. And take the word *thing*, which expresses what seems in a way to be the most basic concept we have. Chomsky points out that some sticks lying on the ground constitute a thing if left there by a human being as a signal; but they are not a thing if left there randomly by a forest fire. Such subtleties abound.

Are any words immune from the kind of complications we have seen? It is sometimes thought that we might be able to give precise definitions for words from technical domains like science or mathematics. But even here things are more complicated than we might like. Take the word *metal*, for example. The following is an excerpt from a lecture on metals by the distinguished metallurgist Robert

Pond. He begins by asking the audience to come up with a definition of *metal*. Their efforts are not successful.

Well, I'll tell you something. You really don't know what a metal is. And there's a big group of people that don't know what a metal is. Do you know what we call them? Metallurgists!... Here's why metallurgists don't know what metal is. We know that a metal is an element that has metallic properties. So we start to enumerate all these properties: electrical conductivity, thermal conductivity, ductility, malleability, strength, high density. Then you say, how many of these properties does an element have to have to classify as a metal? And do you know what? We can't get metallurgists to agree. Some say three properties; some say five properties, six properties. We really don't know. So we just proceed along presuming that we are all talking about the same thing.

Even metallurgists, then, cannot agree on a definition of the word *metal*.

Perhaps *metal* is somehow too broad a term. How about *gold*? Gold is an element of the periodic table and can be pinned down, as it were, with some exactness: it is the element with atomic number 79. What if we define *gold* as 'the element with atomic number 79'? Well, one problem with that suggestion is that most people who know the word *gold* do not know that gold's atomic number is 79. The current suggestion would imply that most competent English speakers do not know the meaning of the word *gold*, which would be a rather paradoxical state of affairs. What is it, we might ask, that allows such people to use the word appropriately? If they do not know its meaning, how is it that they use it quite successfully to talk about gold? A further problem with this suggestion can be brought out with another fantastical scenario. Imagine that an evil demon has been systematically deceiving all the scientists who have ever studied gold. The demon has been making them think that gold's atomic number is 79, but actually it is something else entirely. In fact it turns out, in this scenario, that there is no element with atomic number 79. Now suppose that this remarkable state of affairs is discovered. If the word *gold* just meant nothing other than 'the element with atomic number 79', and if there were no element with atomic number 79, it would seem that scientists would be quite justified in announcing, 'There

is no such thing as gold.' (Compare 'There is no such thing as the element with atomic number 79', which would be quite true under the circumstances.) But in fact, of course, they would not be justified in announcing that. What they would actually say in such a scenario would be something like 'Gold does not have atomic number 79 (but it does exist).'

We could abandon the attempt to define *gold* by means of atomic numbers and concentrate instead on visible characteristics of gold that lay people can appreciate, such as its glittery yellow colour, its ductility, and so on. But this now looks horribly similar to lexicographical attempts to define *chair* by means of number of legs, use for sitting, and so on; and we would not be surprised to find similar difficulties arising. In this case, the existence of fool's gold (iron pyrites) would make it particularly tricky to come up with a definition of this kind that would not include too much. What we would need, of course, would be some means of telling apart gold and fool's gold. How do we do that? Why, we appeal to facts about their chemical make-up such as atomic numbers. But then we are back where we started.

One could object that *metal* and *gold* are words of ordinary language that have been co-opted by science, and that the trouble we have defining them reflects this peculiar status. What about terms that were coined in the course of explicitly theoretical speculation? I am afraid that the prospects of successfully defining words like this are not much better than the prospects of defining *gold*, and for very similar reasons. Take *atom*, for example. Suppose we attempt to give some definition that sums up current thinking about atoms, such as 'unit of matter that consists of a nucleus containing one or more protons (and optionally one or more neutrons) surrounded by a cloud of electrons'. Suppose further that we can dismiss worries about people being competent to use the word but not knowing these details; the word *atom* is sufficiently recondite, we can assume, that anyone who is competent to use it knows at least this much about atoms. It is still possible that some scientific discovery should radically change our conception of atoms, meaning that this definition no longer reflected the best current understanding of them; and yet we would still almost

certainly keep the word and say things like 'Atoms are not units of matter that consist of a nucleus containing one or more protons (and optionally one or more neutrons) surrounded by a cloud of electrons after all.' This is evidence that the meaning of the word *atom* is not the definition just given, or anything along similar lines; for, if it were, it would make more sense to say, 'Since atoms, by definition, are just supposed to be units of matter of the kind we have described, and since we have just discovered that there are no such units of matter, we can deduce that atoms do not exist.'

In the case of the word *atom*, this kind of wholesale revision is not just a hypothetical scenario. The English word derives from the Ancient Greek word *atomos*, which meant 'uncuttable' or 'indivisible'. Atomists, from Leucippus and Democritus in the fifth century BC down to many scientists in the nineteenth century AD, believed that there were ultimate, indivisible units of matter out of which everything else was composed. By the latter half of the nineteenth century, some particular units of matter, called *atoms*, were tentatively identified as being these ultimate, indivisible units of matter. Then came the demonstration in 1897, by the English physicist J.J. Thomson (1856–1940), that these things in fact contained smaller particles, called electrons. What happened? Scientists did not in general conclude, 'These things are not atoms after all, since they are not indivisible.' They said, in effect, 'Atoms are not indivisible after all.' So the word *atom* did not mean 'ultimate, indivisible unit of matter'.

I know of only one area where it seems likely that we have good definitions of words: mathematics. I can see nothing wrong, for example, with the statement that *prime* means 'integer greater than one that has no factors other than itself and one'. It is a matter of some intellectual interest why mathematical terms should be immune from the general chaos that surrounds definitions; but I will not attempt to address this question here.

It is appropriate, at this point, to step back and reflect on what these examples show us. All I have been trying to demonstrate is that giving definitions of words is a task of mind-boggling complexity; by reporting on the state of the art in fields such as epistemology and metallurgy, I have been trying to suggest, but not to demonstrate

conclusively, that no-one has ever given an adequate definition of a word, as far as we know, with the possible exception of mathematical terms; and by discussing a couple of dictionary entries, I hope to have convinced you that dictionary entries do not generally give the meanings of words. Some of these conclusions may be surprising, if you have never studied semantics before. But it is important to realize that they are also rather limited.

To start with, the fact that it is astonishingly difficult to give definitions of words does not show that it is impossible. Even the conclusion, if I could establish it, that no-one has ever given an adequate definition of a word would not show that. Perhaps we just have to try harder and eventually we will hit on some good definitions. Or perhaps definitions of words could in principle be given—perhaps a hyper-intelligent alien race could give some, for example—but human beings are just not smart enough to do this. This last possibility, although it might, once more, strike some readers as surprising, is really not very radical. Imagine trying to explain the atomism of Democritus, or the cathode ray experiments of J.J. Thomson, to a cow. However much you explain atomism, the cow is just not going to get it. Various thinkers have pointed out that some topics could stand in the relation to us that atomism and cathode ray experiments stand in to the cow: we are just too deeply stupid to grasp them. Maybe accurate definitions of words constitute one such topic.

So much for the question of whether we can give definitions of words. But we should also address the question of what definitions of words actually are (or would be, if we could give any). In particular, if we had a completely successful definition of a word, would it be the meaning of that word? Not in a sense that would ultimately satisfy us. The problem is that when we give a scientific or philosophical account of something, we ideally want to explain the thing in question in terms of other kinds of things, things that we take to be somehow more basic. A chemist explains water as a compound of hydrogen and oxygen; a physicist explains atoms as structures involving protons, neutrons, and electrons; a philosopher explains knowledge as true, justified belief of a certain kind. (We wave our hands a little during the last few words of that sentence.) But a

definition is just a string of words. It is unsatisfying, therefore, to say that the meaning of a word is a definition, because that would be to say that the meaning of a word is just more words. It would appear that we were not progressing to any explanatorily deeper level. This is not to say, however, that effort put into constructing definitions is just wasted. As we have seen, efforts of this kind can turn up intricate and sometimes surprising facts about meaning; and any theory of meaning that purported to tell us what meanings were would also ultimately have to account for these facts.

So what things could the meanings of words be? I turn to this topic in the next chapter.

2

What are word meanings?

What things could the meanings of words be? This is no place for a history of the topic. So I will concentrate for the purposes of this chapter on two ideas that have been prominent in the specialist literature for the last few decades: the referential theory of meaning and the internalist theory of meaning.

These theories make use of some occasionally unintuitive concepts from contemporary philosophy, linguistics, and psychology. Among other things, we will read in this chapter that some linguists think that languages like English and Spanish do not exist; and that some philosophers think that Santa Claus does exist. So here, to start with, is a brief overview to establish the lie of the land.

One of the main facts about language that theorists have to account for is that people are able to use it to talk about the world. Even though we can do other things with language, like write nonsense verse, a lot of what we do with it, and lot of its utility, consists in this apparent connection with the things around us. We can describe the world, ask what it is like, and even order parts of it (sentient parts, preferably) to abide by our will. The debate in this chapter focuses on the nature of this word-world relationship.

Both sides of the debate take for granted that the meanings of words are what enable them to hook up with the world. (It is not the pronunciation, for example, that is responsible.) The *referential theory of meaning* proposes the most direct mechanism: meanings of words simply are things in the world. So the word *Iceland*, for example, has as its meaning that very island, a huge chunk of rock and ice in the northern Atlantic Ocean. So once you have grasped the meaning of the word *Iceland*, you automatically know what that

WHAT ARE WORD MEANINGS?

word picks out in the world; for the meaning of the word just is what it picks out in the world. Now this may be all well and good with a proper name, like *Iceland*. But what of other types of words? What does the word *icy* pick out, for example? Or the word *the*? The chief problem faced by the referential theory, as we will see shortly, is that in order to provide things for all these words to pick out, theorists have to posit the existence of increasingly bizarre entities in the world, including, as I said, Santa Claus and a slew of other personages whom sober reflection had previously consigned to the status of myth.

On the other side of the floor, we have the advocates of the *internalist theory of meaning*. They suggest that word meanings are most fruitfully thought of as ideas or concepts in our heads. Take a concept, such as the concept I have of Iceland. It is some psychological entity. Ultimately, if we are correct to suppose that we do our thinking with our brains, this concept of mine is presumably a structure composed out of cells inside my head. (How all this works in detail, of course, is the profoundest of mysteries.) Since the island of Iceland resembles or *falls under* this concept of mine, I use this concept to think about Iceland. And since the concept also forms part of a word (i.e. since it is the meaning of a word), I use that word, *Iceland*, to talk about Iceland. Inside your head, you presumably have a very similar concept that forms part of a very similar word, so that when you hear me say 'Iceland' your concept of Iceland is activated. As you may have noticed, this internalist way of looking at things implies that we each have our own word *Iceland*, and cannot rule out the possibility that the associated concepts are significantly different. Not everyone is happy with this, since it seems to allow drastic failures of communication.

That is the debate in a nutshell. The details that follow can appropriately be thought of as fleshing out this summary.

Let us return to the referential theory of meaning. It says that the meanings of words are things in the world, most of which are not in our heads. Meanings, according to this view, are *referents*, or things that are picked out or referred to. It is perhaps easiest to illustrate the idea with proper names. Consider a proper name such as *Elizabeth II*.

The referential theory of meaning says that the meaning of the proper name *Elizabeth II* is Elizabeth II, Defender of the Faith and Queen of Great Britain, Northern Ireland, and the British Dominions beyond the Seas. The idea is that when you say something like *Elizabeth II is wise*, the name *Elizabeth II* contributes that very woman to what you are saying; and then you say of her that she is wise. But what about the phrase *is wise* and other predicates (like *icy*)? Is there any way to find referents for them? Why, yes. The idea here is that the word *wise* refers to the *property* of being wise.

At one level, the notion of a property is familiar and homely enough. A property is an aspect or characteristic of something. This apple has the property of being red; that one has the property of being green; both have the property of being an apple. But already, in saying this much, we are entering upon controversial philosophical ground. Some philosophers have claimed that, if my red apple and my green apple both have the property of being an apple, then this implies that there is a separate thing, the property of being an apple, which both of my apples have. A property of the philosophical kind would be rather similar to a property of the commercial kind, in that both kinds of thing can be jointly possessed. Many philosophers have thought this, beginning with Plato, whose *Forms* are often taken to be forerunners of properties in this sense. Plato would have said that my red apple and my green apple separately instantiate the Form of the Apple, which is an eternal, unchanging, non-physical entity that is something like a blueprint for all earthly apples; only by instantiating the Form of the Apple can any object be an apple. Many philosophers since Plato, in ancient, medieval, and modern times, have held similar views: there is a property of being an apple, which all actual apples instantiate.

The property, in this way of thinking, is a *universal*, in that it is present simultaneously in numerous different objects. The opposite of a universal is a *particular*, which is just an ordinary object that is not present in different places at the same time; but the theory that is relevant to us holds that properties are universals.

Theories along these lines have been invoked to account for various phenomena. One example is the nature of similarity. What is

it for two things to be similar? It is for them to instantiate at least one property in common, according to this view. Most notably for our present purposes, this kind of theory has been appealed to by Plato and many successors in order to explain the meaning of what are called *general terms*. A general term is a word that is applicable to more than one thing, like *apple*, *runs*, *wise*, *icy*, and other nouns, verbs, and adjectives. The idea is that general terms stand for properties.

Properties are often paired with *relations*. Relations, according to the relevant theories, are also universals, but instead of being instantiated by just one object, wherever they are present, they are supposed to hold between two or more objects. Seeing, for example, might be a relation; it holds between the seer and the seen, and would be the referent of the verb *see* according to the referential theory of meaning. We will see this idea spelt out in Chapter 6.

A neat formal way of summing up the foregoing ideas on meaning that is quite popular in the philosophical literature is the *Russellian proposition*, invented at the beginning of the Twentieth Century by the philosopher Bertrand Russell (1872–1970), one of the founders of analytic philosophy. A *proposition* is the meaning of a declarative sentence. Russell would have claimed that the meaning of the simple declarative sentence *Elizabeth II is wise* was the ordered pair (Elizabeth II, wisdom). The notion of an *ordered pair*, indicated by *angle brackets* ($\langle \dots \rangle$), derives from set theory; an ordered pair is a set of two things that has a first member and a second member. The first member in this example is Elizabeth II (the monarch, not the name) and the second member is wisdom, or the property of being wise. A sentence whose meaning is a simple Russellian proposition like this would be true if and only if the first member of the proposition has or instantiates the second member.

Since set theory will recur in this book, here is a brief reintroduction to it for those whose schoolroom memories of the subject are hazy. Others can allow their attention to wander freely for the duration of this paragraph. A *set* is a collection of objects. Sets are commonly written by means of curly brackets: the set whose only members are the numbers one and two would be written $\{1, 2\}$

or $\{2, 1\}$. There is no ordering imposed on the members of ordinary sets like this, and the two expressions just written down represent the same set. Things are different with ordered pairs, however, which are the special kinds of sets just described: $\langle 1, 2 \rangle$ is not the same set as $\langle 2, 1 \rangle$, since the order that the elements appear in is significant. That is enough set theory for now. More will be introduced along the way.

Reference can fruitfully be thought of as the contribution of objects, whether ordinary ones or properties, to Russellian propositions. (By the way, let no-one suspect me of *lèse-majesté* in letting the word *object* be applicable to Elizabeth II—I am using it in the philosophical sense, whereby people are objects too.) The referential theory of meaning arguably has an advantage over the attempt to capture meaning by definitions. To return to our former example, suppose that I have a favourite chair that I call Albert. The sentence *Albert is a chair*, then, would be associated with the Russellian proposition $\langle \text{Albert}, \text{chairhood} \rangle$, where *chairhood* is the property of being a chair. (To keep things simple, I am abstracting away from any contributions that might be made by *is* and *a*.) If there really is such a thing as the property of being a chair, it seems that we might have found the meaning of the word *chair* right there, with none of the complicated speculation about leg-count and similar matters that characterized our earlier discussion.

You might be thinking that this sounds a bit too easy. Can we really just decide that one of the ultimate constituents of the world is the property of chairhood and designate this novel entity, with no further ado, as the meaning of the word *chair*? And indeed there is a long and honourable philosophical tradition of scepticism concerning properties, and concerning universals in general. A *realist* concerning universals is someone who accepts their existence, and many eminent philosophers, with Plato at their head, have been realists; but opposed to this camp are the *nominalists*, who believe that there are no such things as universals. This tradition seems to have got under way in the middle ages, with such distinguished exponents as William of Ockham (c.1287–1347); and it continues to the present day. William of Ockham, it is good to recall, is the man who gave his name to Ockham's Razor, the maxim that enjoins theory-builders

of all kinds not to admit more entities into their theory than is absolutely necessary. Is it really necessary to have our ontological theory accommodate these very bizarre entities that can be in lots of different places at once? Fascinating as the debate is, it would take us too far afield to go into it now; suffice it say that there is no consensus on whether properties of the kind we have been talking about really exist. So they do not obviously provide a firm foundation for the semantics of general terms.

A similar point can be made with regard to proper names, which might perhaps have seemed to constitute an area where the theory performed well. There are lots of proper names, such as *Sherlock Holmes* and *Santa Claus*, that do not name anything that actually exists. But if Santa Claus does not exist, the name *Santa Claus* cannot refer to anything. And yet *Santa Claus* is perfectly meaningful and capable of being used in lots of perfectly innocuous sentences, such as *Santa Claus does not exist*. And so it seems that the referential theory of meaning has run into another problem.

Some advocates of the referential theory, in response to this problem, have seriously proposed that Santa Claus does in fact exist. I assure you that I am not making this up. To be fair to these theorists, they are not claiming that at Christmas we really are liable to have a reindeer-borne, red-clad man crawl down our chimneys. They maintain that Santa Claus is an *abstract object*. In order to assess this claim, we will naturally want to know what abstract objects are supposed to be. But here, perhaps predictably, we encounter further complications.

The difference between abstract and *concrete* objects is widely thought to be of fundamental philosophical significance. All objects are sometimes claimed to fall into one of these two categories. Some paradigm examples of concrete objects are you, me, my computer, the desk at which I am sitting, and my copy of P.G. Wodehouse's *Jeeves and the Feudal Spirit*. Some paradigm examples of abstract objects are numbers, geometrical shapes, and P.G. Wodehouse's *Jeeves and the Feudal Spirit*. (The geometrical shapes of which I speak are the perfect triangles, circles, and so on, studied by mathematicians; and the distinction I am aiming at with the example of the

Jeeves novel is the distinction between the content of a book, which can be manifested in many individual copies, and the individual copies themselves.) The standard conception of abstract objects is that, in contradistinction to concrete objects, they are not located in space or time and do not engage in causal relationships. This makes a certain amount of sense when gauged against the examples I just mentioned. Concrete objects like you and my desk occupy certain locations in space and typically come into and go out of existence. (If you think that there is an eternal afterlife, so that you will never go out of existence, you probably believe, nevertheless, that you came into existence.) The number two, however, and a perfect circle with a radius of exactly 2 cm are not the kinds of things that you will ever pinpoint on a map or trip over on the pavement. (You might see things that approximate to being perfect circles, but it is unlikely in the extreme that you will ever find an absolutely perfect circle in the sublunary world: the circumference of a perfect circle has no width, for one thing.) And, although certain knowledge of these matters may be hard to come by, it seems unlikely that the number two came into existence at a certain point in time or will ever cease to exist. Nor, arguably, can numbers or geometrical shapes cause things to happen.

It is perhaps already evident that works of fiction depart from this standard conception of abstract objects in some ways. In particular, it seems that they come into existence at certain times: they are created by authors. (The alternative is that *Jeeves and the Feudal Spirit* has always existed, at least since the beginning of time, if time had a beginning.) The same thing will presumably apply to fictional characters and characters from folklore: if Jeeves and Sherlock Holmes and Santa Claus are abstract objects, they will presumably be rather unusual ones in that they came into existence at particular times and were created by human beings. But there is nothing to prevent the advocates of the position that Santa Claus exists and is an abstract object from saying that there are different types of abstract object: some, like numbers and geometrical shapes, did not come into existence at particular times, whereas others, such as Santa Claus, did; all, presumably, are not located in space and do not participate in causal relationships in the same way that rocks and desks do.

This theory of fictional characters as abstract objects must be the approximate content of an important part of the referential theory of meaning. What is its status? It turns out to be rather similar to the theory's use of properties. Although many philosophers believe in abstract objects, there are many who do not. The view that there are no abstract objects is, rather confusingly, called *nominalism*. We must distinguish between nominalism about universals, which we saw above, and nominalism about abstract objects; the two are clearly conceptually distinct, and each view can be held consistently without the other. Meanwhile, the view that in fact there are abstract objects is called *platonism*, also rather confusingly. This latter term is rather confusing because it is doubtful that Plato himself was an advocate of platonism in this sense—although his Forms are eternal and are presumably not located in space, they nevertheless have strong causal powers. But these terminological matters aside, why might one be a platonist? Think of numbers, a platonist might reply. If it is true that $2 + 3 = 5$ (and one is hard pressed to deny it), then numbers must exist, for how could this statement about the numbers two, three, and five be true if these numbers do not even exist? But if numbers exist, then we have abstract objects, for surely, as I said above, we are not going to pin down the location of the number two in space or time; nor is there any reason to think that it participates in causal relationships. Why, alternatively, might one be a nominalist about abstract objects? Ockham's Razor, a nominalist will urge, dictates that if we can do without these exceedingly weird entities then we should; and besides, if abstract objects are not spatiotemporally located and do not participate in causal relationships, how do platonists suppose that anyone can acquire knowledge of them? For our acquiring knowledge about something is an event and, by definition, abstract objects do not participate in causing events; so any knowledge that we think we have about abstract objects is either pure confusion or, at best, a dim and confused apprehension of other kinds of objects. This objection, sometimes called the *epistemological argument against platonism*, is a stubborn one. Again, as with the case of nominalism about universals, it would be impracticable to examine in detail all the arguments to be found in this neck of the

philosophical woods; suffice it to say that it is by no means obvious that there are any such things as abstract objects, or that we could know about them if there were. It is not clear, then, that it is legitimate to appeal to abstract objects in defence of the referential theory of meaning.

A related concern is the following. The referential theory of meaning says that Santa Claus exists and is an abstract object; furthermore (and this was the point of this metaphysical excursus), the proper name *Santa Claus* is provided by this doctrine with a referent, which it must have if the referential theory of meaning is to be true. So *Santa Claus* refers to a particular existent abstract object. But then the sentence *Santa Claus does not exist* is predicted to say of a particular existent abstract object that it does not exist, which must, of course, be false. But intuitively *Santa Claus does not exist* is true. And this sentence does not seem, intuitively, to be denying the existence of an abstract object; it seems to be denying the existence of a putative concrete object, a jolly, red-clad man who is liable to climb down our chimneys at Christmas. Even if we grant referential theorists the existence of Santa Claus (which is surely about the most generous intellectual concession that one could ever make), problems still abound here. And when the non-existence of Santa Claus poses a problem for your theory, it is time to look for a new theory.

Let us move on, then, to examine the internalist theory. This theory states that word meanings are internal mental structures. The basic idea can also be expressed by saying that meanings are ideas or concepts; or by saying that meanings are in the head. Some version of the theory goes back at least to Aristotle (384–322 BC), who in his treatise *On Interpretation* wrote, ‘Spoken words are symbols of mental experiences.’ The theory was prominent in the middle ages: the fourteenth-century French philosopher John Buridan, for example, in his *Summulae de dialectica*, declares that ‘the capability of speaking was given to us in order that we could signify our concepts to others and also the capacity of hearing was given to us in order that the concepts of speakers could be signified to us.’ In the early modern period, a prominent exponent of this view was the English philosopher John Locke (1632–1704). In his *An Essay Concerning*

Human Understanding (1690), Locke wrote, ‘words, in their primary or immediate signification, stand for nothing but the ideas in the mind of him that uses them.’ And in modern times the view is held by Noam Chomsky and many other linguists and is standard in psychology.

Here is a Chomskyan version of the internalist theory. Language in general, for Chomsky, is basically a psychological phenomenon. Human beings are equipped with certain specialized mental apparatus such as a *mental lexicon*, which contains all the words we know, and a syntactic module, which tells us how to arrange words in grammatical sentences. The whole ensemble of an individual’s language-specific mental apparatus is called that person’s *language faculty*. Words, according to Chomsky, are mental entities that consist of three parts: phonological information, which tells us how to pronounce them; syntactic information, which tells us what part of speech they are and such things as whether they obligatorily take a direct object (in the case of verbs); and semantic information, or meaning. (In the case of literate people, we can add orthographical information to this list.) The semantic information must be very intricate, since it must give rise to all the phenomena that we noted above when trying to give definitions of words; but it is nearly all inaccessible to consciousness, and details of it can be reconstructed only with painstaking effort, as we have seen. As for the precise form that this semantic information takes, Chomsky has little to say; it is deeply mysterious. (We will see in a short while that contemporary psychologists have tried to flesh out the picture a little.)

Before we go on, it is worth pointing out some consequences of this view. One is that our talk of things like ‘the English word *chair*’ is misleading. Strictly speaking, according to the internalist view, it is not the case that there is one word *chair*. We should rather say that in your mental lexicon there is a word *chair* with certain phonological, syntactic, and semantic features and in my mental lexicon there is another word *chair* with features that are similar but possibly slightly different. To be clear about this, we should introduce some philosophical terminology about difference. In the case of *numerical difference*, there are two separate objects residing in different places.

In the case of *numerical identity*, we are concerned with exactly one object: we might say that Superman and Clark Kent are (numerically) identical, which boils down to *Superman* and *Clark Kent* being two different names for the same man. Numerical difference (or identity) is to be contrasted with *qualitative difference* (or identity or similarity). Two things could be numerically distinct but qualitatively very similar: two new cars of the same make and colour, for example, would be qualitatively very similar but numerically distinct. The case of the word *chair* in your head and the word *chair* in my head, then, is definitely a case of numerical difference, in that two separate objects are involved; and it may very well be a case of qualitative difference too, in that the chance of our pronouncing these words exactly alike is not very high. (A trained phonetician could almost certainly find some small differences.) It turns out that there is also a good chance that your word *chair* and my word *chair* will differ very slightly in semantic features too, as we will see later. Extrapolating from the case of single words, Chomsky also maintains that there is, strictly speaking, no such thing as English, or French, or Japanese, or any other natural language. If language is purely psychological, and individual speakers' language faculties are the only linguistic things there are, there is no place for any separate object 'English'. There are just lots of groups of human beings with language faculties that resemble each other in sufficient detail for communication to be able to take place; one of these groups we informally call 'English-speakers', another one 'French-speakers', and so on.

So for Chomsky word meanings are parts of words and are all in the head. Now it might seem as if there is an obvious and grave problem with this kind of view, which was pointed out by the English philosopher John Stuart Mill (1806–1873) in his *System of Logic* (1843): "When I say, "the sun is the cause of day," I do not mean that my idea of the sun causes or excites in me the idea of day; or in other words, that thinking of the sun makes me think of day.' The difficulty is that, if word meanings are just ideas in our heads, it is not obvious how we can use words to talk about things other than ideas in our heads. How do we talk about the outside world? (Note that the referential theory of meaning, whatever may be its flaws,

solves this problem in a very immediate way, since word meanings, according to this theory, actually are things in the world.) According to Chomsky, the crucial notion here is *use*: we use words rather like tools to talk about things in the world or focus attention on particular aspects of the world. Responding to the American philosopher Hilary Putnam, who was insisting that there is some relation between speakers, words, and things in the world, Chomsky wrote, 'So there sometimes is . . . , in more or less the sense in which a relation holds of people, hands, and rocks, in that I can use my hand to pick up a rock.' Our usage and understanding of words in this way is spontaneous and unthinking. There must be some deep-seated convention, then, perhaps instigated by biology rather than society, to the effect that when people utter a word whose meaning is a particular internal concept, they are not attempting to draw attention to the concept itself, but rather to things that, as we say, *fall under* that concept.

One advantage of the internalist theory that should already be evident is that it deals much more happily than the referential theory of meaning with terms like *Santa Claus*. Everyone who has heard of Santa Claus presumably has a Santa Claus concept in their heads; they know that nothing, in fact, falls under this concept, but that does not matter. In order to have a meaning for the proper name *Santa Claus*, all we need is the concept.

The internalist theory also deals more readily than the referential theory with another intriguing fact about meaning. In various works, Chomsky has drawn attention to sentences like *Jeeves and the Feudal Spirit is a best-seller and weighs twelve ounces*. What exactly is it that is a best-seller and weighs twelve ounces? I might be tempted to brandish my copy of the book and say, "This is." My particular copy might weigh twelve ounces, certainly. But my particular copy cannot be best-selling all by itself: many different copies of a book must sell in order for it to be best-selling. What kind of thing can be best-selling, then? In the light of our previous discussion, the obvious candidate is an abstract object: recall that some philosophers claim that the content of a book is an abstract object that might be realized, as it were, in many concrete copies. So an abstract object might possibly be best-selling, if we follow these philosophers. But

an abstract object cannot weigh twelve ounces: since abstract objects are not located in space, they cannot be affected by gravity, which is necessary in order to weigh anything. It seems, then, upon sober reflection, that no actual object can both be a best-seller and weigh twelve ounces. Chomsky sums up the conundrum by saying that the meanings of words of natural languages evidently presuppose the existence of objects that are simultaneously abstract and concrete, which is an impossibility for any actual object. But this presents the referential theory of meaning with a grave difficulty when it comes to finding a referent for the proper name *Jeeves and the Feudal Spirit*, as used in the sentence above: it seems there can be no such object; but the referential theory claims that the meanings of words are objects. The internalist theory, on the other hand, has less of a problem here: it just has to point out what everyone already knows all too well, namely that ideas can be self-contradictory.

This is not to say that there is no problem at all here for the internalist theory of meaning. We still have to explain how it is that we can say something apparently straightforward and true, like the above example, by using self-contradictory concepts. I am not aware of any work on this.

Let us move on to examine the account of word meaning given by contemporary psychology. It seems beyond doubt that normal human beings are equipped with *concepts*. Concepts are mental representations that allow us to classify things we come across and access memorized information about them so that we will know how to behave appropriately towards them. If this object in front of me falls under my concept CHAIR, for example, I know that I can sit on it (unless it also falls under my concept TOY or my concept ART EXHIBIT). If, however, it falls under my concept TIGER, I will not try to sit on it, unless I am possessed of very unusual skills. There is some impressive evidence from experiments on concepts that seems to indicate that the meanings of words (or at least nouns and verbs) are concepts. For example, concepts display a variety of distinctive properties, one of which is *typicality effects*: there seem to be good and less good members of the set of things that fall under a given concept, as diagnosed by how easily experimental subjects can clas-

sify them. So if experimental subjects are exposed one at a time to photographs of various objects and instructed to press one button if the object is a bird and a different button if the object is not a bird, with their responses being timed, they will take longer to verify that an emu or a penguin is a bird than they will to verify that a sparrow or a wren is a bird. The idea is that sparrows and wrens in some way correspond more closely to the concept BIRD than emus and penguins do, even though the latter do indeed fall under this concept. Now the point is that in this kind of experiment exactly the same results are obtained whether the stimuli are photographs or words. So if the previous experiment is repeated exactly, with the exception that words and not photographs are flashed up on the screen, subjects will take longer to verify that the words *emu* or *penguin* name birds than they will to verify that the words *sparrow* or *wren* name birds. The logical conclusion is that the same information is being drawn upon each time: there seems to be one concept BIRD, for example, that is accessed no matter whether we are shown a picture of a bird or shown the word for a kind of bird. Since the concept BIRD naturally contains a lot of information about birds it would seem to be a good candidate for the meaning of the word *bird*, and any theory that postulated both the concept BIRD and a separate word-meaning would be seriously uneconomical. Psychologists, therefore, suppose that word meanings are concepts, and that the meanings of complex phrases and sentences are concepts too, derived by combining the constituent concepts.

What theories do psychologists propose about the structure of concepts? There are several theories on offer. For current purposes, it will suffice to examine the best known one—the *prototype theory*. The prototype theory was proposed by the American psychologist Eleanor Rosch in the 1970s. Imagine a concept and some things that fall under it—the concept BIRD and various birds, for example. The prototype theory says that the concept is a summary representation of features that the things in question can have, together with weightings indicating how important it is to have those features in order to fall under the concept. For example, in the case of BIRD, there might be a feature ‘feathered’ that would be highly weighted, since birds

quite generally have feathers. (Not plucked ones, though.) Similarly for features like 'egg-laying', 'winged' and 'capable of flying'. But other features like 'brown' would be less highly weighted. In fact, since birds can be all kinds of different colours, lots of contradictory colour features would be listed, each with fairly low weightings. So what happens if we are presented with an object and asked if it is a bird? Basically, we go through the features in the concept BIRD and add up the weightings of the features that the object has; we might also subtract the weightings of at least some features in the concept that the object does not have. If the final sum exceeds a certain threshold, the *categorization criterion*, we judge that the object falls under the concept—for example, that Tweety is a bird. Otherwise, we judge that the object does not fall under the concept. It can be seen how this theory accounts for the typicality effects that we just noted. In the case of a sparrow or a wren, all of the heavily weighted features ('feathered', 'winged', 'beaked', and so on) are present, and they easily combine to exceed the categorization criterion. But in the case of a penguin, many of these heavily weighted features are not present, or are not obviously present. (Do penguins have feathers? It turns out that they do, but from a distance they look rather like they have fur. And do those flippers count as wings? They certainly do not enable their owners to fly.) So the categorization system is left scrabbling, as it were, to exceed the threshold by means of less heavily weighted features, or by heavily weighted features only belatedly recognized as being applicable.

While its treatment of typicality effects is laudable, the prototype theory is less promising in other arenas. One important example involves *compositionality*. The principle of compositionality states that the meaning of a complex phrase is determined solely by the meanings of its parts and their syntactic arrangement. Some principle of this sort seems necessary because of the *productivity* of language: we can produce and understand an indefinitely large number of novel phrases whose meanings cannot be listed separately (phrase by phrase) in our mental lexicons. So, for example, you understand me if I write about a *winged chihuahua*, *Santa Claus on a Harley Davidson*, and a *dramatic reading of the telephone directory*, even if

(as is likely) you have never heard or read those phrases before. We must be able to take the words out of which these novel phrases are constructed and combine their meanings somehow in order to arrive at the meaning of the whole. It has been alleged by the American philosopher Jerry Fodor that the prototype theory fails to account for compositionality in a range of basic cases. Take the phrase *pet fish*, for example. Its meaning, according to the prototype theory, must be a concept of the kind described above. So it must give heavy weighting to features that pet fish generally have, such as 'brightly coloured', 'small', and 'lives in bowls or small tanks'. But it is hard to see how a concept with these features can be the result of combining the concepts attached to the words *pet* and *fish*: neither pets in general nor fish in general are brightly coloured, small, or dwellers in bowls or tanks. So the meaning that the prototype theory would give to *pet fish* cannot be derived systematically from the meanings that this theory would give to *pet* and *fish*. This is generally seen as a bad thing.

There are other theories of concepts in psychology, but they are subject to other problems. This is not to counsel despair. A tremendous amount of progress has been made in the psychology of concepts since the 1970s. The discovery of typicality effects, for example, is a major landmark that must be accounted for by any theory in this area. And for our current purposes, we do not need to profess allegiance to any particular theory of concepts in order to hold the view that word meanings are concepts, if by that we just mean that word meanings are internal mental structures.

One objection that seemingly threatens every internalist theory of word meaning focuses on the apparent possibility of accurate communication. This objection goes back at least to the great German philosopher Gottlob Frege (1848–1925), the founder of modern logic and analytic philosophy. If word meanings are internal mental structures, then it is entirely possible that you and I have different meanings attached to what we ordinarily take to be the same word. Take the word *jejune*, for example, whose traditional definition is something like 'meagre, unsatisfying'. In one sense, of course, the internalist theory is explicitly committed to the meanings that we attribute to *jejune* being different: they are numerically distinct objects. Now

Frege and many philosophers since have seen sinister implications in the idea that a word has lots of numerically distinct meanings, a different one in the head of every person who knows it. For what could guarantee that these numerically distinct meanings would be qualitatively identical? And if the numerically distinct meanings are not qualitatively identical, then it seems that grave breakdowns of communication could ensue. You would mean one thing by *jejune* and I would mean something significantly different. We would be talking past each other.

It is certainly true that there is an impressive amount of qualitative similarity between the meanings that people seem to assume for a given word; and it is true that linguistic communication often proceeds quite well. But it is far from clear that we need to fall back on the referential theory of meaning in order to ensure this. The trouble is that the referential theory, even though it says that meanings themselves are things in the world, must provide some account of how people get to know about these things; for people must have mental representations of word meanings in order to function linguistically, even if word meanings are not mental representations. What would the meaning of the word *jejune* be according to the referential theory? Presumably it would be the property of jejuneness. Now how likely is it that people will be able to get a clear idea of that property? And how likely is it that everyone will converge on exactly the same idea? Our previous discussions come back into play here. To start with, some philosophers believe that properties are abstract objects. But that, as we have seen, runs the risk of making them immune from any means we might have of gaining knowledge about them. What about the other option, which is to say that jejuneness is a universal and a concrete object? Then it has to be an extremely strange concrete object that is present in a spatially disconnected fashion in everything that is *jejune*. How likely are people to gain a good idea of something as complicated as the universal of jejuneness? And with what senses or mental faculties do they do so? It is entirely unclear how this is supposed to work. It is somewhat mysterious, then, how appealing to the referential theory of meaning is supposed to help us at this point.

Furthermore, when we come to examine the alleged interpersonal identity of meaning that the referential view is supposed to guarantee, and the internalist view imperil, it turns out to be a chimera. To start with *jejune*, many people who know this word associate it with a meaning something like 'puerile', as if it were related to French *jeune* 'young'; it is occasionally even spelled *jejeune*. Others, as I mentioned, associate it with a meaning 'meagre, unsatisfying', which is derived from that of the ancestor of the word, Latin *jejunus* 'fasting'; people who use this meaning of the word are likely to castigate the 'puerile' meaning as a malapropism, but it cannot be denied that both meanings exist, in the internalist sense. So here is a first example of lack of interpersonal identity of meaning. And we do not need to search out relatively rare words like *jejune* to make this point. The psychological literature reveals substantial speaker variation even in the case of common words. In a classic 1978 paper, the Princeton psychologists Michael McCloskey and Sam Glucksberg tested thirty Princeton undergraduates twice, in sessions one month apart, on whether they thought that items designated by given words were members of categories indicated by further words. Subjects were shown pairs of words like *handkerchief-clothing*, for example, and asked to indicate whether things that could be described by the first word could also be described by the second. It was found that 45% of responses indicated that a handkerchief was an item of clothing, while 55% indicated that it was not. This indicates quite substantial disagreement over the meaning of *clothing*, *handkerchief*, or both. In other findings, 30% of responses indicated that curtains were furniture, while 70% did not; 47% of responses claimed that lobsters were fish, while 53% did not; and, rather alarmingly, it was averred 35% of the time that poets were not animals. You may be relieved to know that women were claimed to lie outside the animal kingdom only 3% of the time. As if this were not enough to subvert the doctrine of interpersonal identity of meaning, the psychologists found that substantial numbers of subjects actually changed the answers they gave between the two sessions. Ten undergraduates changed their minds over the course of the month on whether curtains were furniture, for example; and two had second thoughts on the question of whether

women were animals. (It is unfortunately not recorded which aspect of Princeton life it was that prompted these students to reconsider.) So far are we from the alleged interpersonal identity of meaning that we even have to deal with breakdowns of intrapersonal identity of meaning.

It is time to bring this discussion of the referential and internalist theories of word meaning to a close, at least temporarily. The reader will doubtless have noted my own sympathies for the internalist camp; but I cannot stress too strongly that in this introductory treatment I have tried only to give a flavour of the debate. I have not tried to cover every argument that has been made on this topic; and many philosophers and linguists far more venerable and learned than I am are convinced of the referentialist conclusion.

3

Semantic properties of words

In this chapter I will discuss some semantic properties that words are traditionally supposed to have. I will concentrate on synonymy, ambiguity, and vagueness.

With all the fine variation in the meanings that different people attribute to what we loosely call the same word, you may be wondering whether two different words ever have the same meaning. In other words, you might be wondering about the status of *synonymy*. For two words to be synonymous, traditionally speaking, is for them to have the same meaning. In internalist terms, we would recognize two ways in which this might happen: the meanings that the two words had could be numerically identical or merely qualitatively identical. The first possibility would be realized if one concept in someone's head was simultaneously part of two different words. The second possibility would be realized if two words, whether in one head or in different ones, had qualitatively identical meanings.

But do any synonymous pairs of words really exist? As soon as one starts looking into fine shades of meaning, it might appear doubtful. For example, a patient might lie on a psychoanalyst's *couch* with perfect propriety; but if they were to lie on the same person's *sofa*, we might suspect a breach of professional ethics, since the latter furnishing would probably be in the psychoanalyst's home. Similarly, having a *big* brother has chronological implications that having a *large* brother does not—to say nothing of the possible totalitarian overtones. But one can push scepticism about synonymy too far. It is sometimes claimed, for example, that pairs of words like *napkin* and *serviette* (or *pudding* and *dessert*) have different meanings merely because they are used by different social classes—the Mitfordesque