

REFERENCE AND DESCRIPTION REVISITEDⁱ

Frank Jackson

Many today speak of the description theory of reference in terms suitable for the grand old, refuted theories of the past: it belongs with behaviourism, phenomenalism, libertarianism, verificationism, and cartesian dualism. I have never understood why. The objections brought against the theory all seem to me either to misunderstand the theory in one way or another, or to overlook obvious candidates to be the descriptions or properties that secure reference according to the theory. To argue this in detail is the work of a book, not an article. I will outline how to defend a description theory of reference for spoken and written language, giving, I hope, enough detail to show what the book might look like. The paper has a rather negative cast; there is a good deal of listing of well-know objections followed by rebuttal. But that is how it has to be when you are taking on conventional wisdom, and some positive points are made in developing the rebuttals. I should also note that although I am in a substantial minority, there have always been defenders of the description theory, and many of the things I say have been said in one form or another, somewhere or other, by someone or other.ⁱⁱ

I start with some necessary background about language before proceeding to look critically at some of the some of the most influential objections to the description theory.

Language and communication

I often want to communicate my view about how things are to someone else. Perhaps I think that some course of action is dangerous and want to tell you so, or perhaps I have a view about who will best govern the country and hope to sway you to my view by telling you, or perhaps.... There are a number of ways I might seek to achieve my end: by pointing, by holding up pictures, by engaging in charades, or, most simply, if you and I share a language, by uttering words and sentences in the language we share (and know that

we share, and know that we know that we share,...). Language—or much of language, and in any case that aspect of language we will be concerned with—is a convention generated set of physical structures that has as a principal function making it easy to articulate, and in consequence easy to record, transmit in communication, debate the correctness of, and so on, how someone—you, me, the enemy, the ideal observer,...—takes things to be.

I am going to presume this Lockean picture of language. To avoid misunderstanding, though, I should note an unfortunate ambiguity in talk of language having the capacity and principal function of capturing and thereby facilitating the communication of how we take things to be. It might be read as saying, implausibly, that language is principally about states of belief, or it might be read as saying, plausibly, that language is about what states of belief are about, namely, how things are.ⁱⁱⁱ On the first reading, the sentence ‘There are electrons’ makes a claim about what someone believes; on the second, sentences and beliefs alike represent how things are, including that there are electrons, and if you want to make a claim about what someone believes, you instead use a sentence like ‘She believes that there are electrons’.

With this as background, we can see the attraction of the description theory of reference. If we are to use physical structures to give information on how we take things to be, we need associations in the minds of transmitters and receivers of the putative information between the various structures and the various ways things might be. We use flags to give information about deaths of the famous, roadworks, the nationalities of visiting dignitaries, and so on. The system depends on known associations. Flying a flag at half-mast flag would not be much use for telling about the death of someone famous if the association between flying the flag at half-mast and death were a dark secret. In the same way, if we are to use the physical structures known as words to tell about how we take things to be, we must associate various words with various ways things might be; or, as we will put it, we must associate words with properties. Here the term ‘property’ does not signify a universal in the sense that figures in debates over the one and the many, but is simply a short word for a way things might be in the wide sense that includes relational and dispositional ways things might be. This way-things-might-be need not be a particularly unified way things might be. If I tell you that something is a fish or a fowl, I

tell you something about how it is. Hence, being a fish or a fowl is, in our relaxed sense, a property—possessed by all the fish and all the fowl, and by nothing that is not one or the other—despite its evidently disjunctive nature. Also, there need not be only one property associated with a word. There may instead be a number of properties. The word may be one to use when we want to say that there is something with a goodly number of a certain list of properties; the word may be, that is, a cluster term. Finally, I should emphasise that in some cases the associated property will be that of being thought about, that is, being what a thought refers to. I might stipulate that the word ‘Fred’ in my mouth stands simply for the object I am thinking about at the time of uttering the word—this stipulation might be useful in playing charades, for example. In this case, the associated property for the word ‘Fred’ would be the property of being the thing thought about at the time of utterance. This highlights a point we will return to: the description theory of reference I am defending is not a theory of reference in thought. It is a theory of reference for terms in a public language (in the sense that the language tokens are publicly available—they are written or spoken or engraved or...) that presupposes that we can refer in thought. This means that I will be taking no position on whether or not thoughts about objects are nothing other than thoughts that there are things that are thus and so—a matter often linked to the description theory. We can grant that ‘Fred’ in my mouth at *T* refers to whatever has the property of being thought about by me at *t* without committing ourselves to the further view that my thought will be that something is thus and so and the object thought about is simply whatever is thus and so. The description theory of reference for terms in a public language is distinct from the description theory of what it is to be an object of thought.^{iv}

If lots of words are associated with properties—something that follows from the way we use words to give and receive information—it is useful to have a name for the relation between these words and the things that have the properties associated with them. The description theory of reference says, first, that a good name for this relation is ‘reference’, and, second, that the words we use to give and receive information via their association with properties include those known as proper and common nouns. As we might say it, using the pre-analytic or folk term ‘about’: terms like ‘London’, ‘Pluto’,

‘water’, and ‘inert gas’ are used by speakers to talk *about* whatever has the properties they associate with the term in question; or, as philosophers of language might say it, a name *T* used by *S* *refers* to whatever has the properties that *S* associates with *T*.^v This quick statement will be refined and explained as we undertake our sceptical tour of the various objections to the description theory of reference, but will remain our guiding thought.

We now start the sceptical tour.

The passing the buck objection

This objection has been put by Michael Devitt in a number of places. Here is a recent statement of it.

Description theories are *essentially incomplete*. A description theory explains the reference of a word by appealing to the application of descriptions associated with the word. So the theory explains the reference of the word by appealing to the reference of other words. How then is the reference of those other words to be explained? Perhaps we can use description theories to explain their reference too. This process cannot, however, go on forever: There must be some words whose referential properties are not parasitic on those of others.... Description theories pass the buck. But the buck must stop somewhere.^{vi}

This objection misunderstands the theory, or at least misunderstands the theory in the Locke-inspired form that I accept. The description theory explains the reference of a word as that which possesses the *property* or *properties* associated with the word. Just as we said above. It is not an essential part of the theory that we should have words, or ‘other words’, for these properties. Indeed, as far as the theory is concerned, there might be a one word language and the reference of that one word might be to whatever has the property or properties associated with that single word. Perhaps one word languages are impossible,

but if they are, this is an additional thesis in the philosophy of language—not a very plausible one, in my view; it is not part of the description theory of reference. In any case, the important point is that the talk of descriptions applying is to be understood in terms of the possession of properties, not in terms of the application of (other) words in the language in question.

Defenders of the language of thought sometimes say that the problem of reference for public words is essentially the same as the problem of reference for words in mentalese, and it might be suggested that we should rephrase the buck passing objection in terms of the language of thought. The rephrased version might run somewhat as follows. According to the description theory, the reference of *T* in the mouth of *S* is to the thing that has the properties *S* associates with *T*; but for *S* to associate properties with *T* is for *S* to have words in mentalese that refer to those properties. This, the objection might conclude, makes it clear that the problem of reference for words in a public language is simply being handed across to words in the language of thought. The buck is being passed from words in public language to words in mentalese.

For the sake of the argument, I will accept that there is a language of thought. The real problem with the rephrased objection is that the problem of reference for words in a public language is very different from the problem of reference for words in mentalese. The difference derives from the point that Locke put by saying that words are *voluntary signs*.^{vii} A word like ‘water’ might have referred to gold, and would have done so if we had agreed to use the word ‘gold’ in the circumstances we in fact use the word ‘water’ in: which words in a public language refer to which things is in part a matter of the largely implicit conventions of usage we enter into.^{viii} This is crucial to the plausibility of the description theory of reference. It is plausible that we follow the convention of using ‘water’ when confronted with stuff we take to be thus and so, and this is why it is plausible that we use the word for stuff that is thus and so. But this picture only makes sense for words in a public language; it would be a nonsense to suppose that we entered into a convention to use the words of mentalese in certain circumstances. Because we do not know what they are, we cannot make agreements concerning them. Moreover, even if we did know what they are, we could not choose when to use them in accord with a

convention of usage; changing when we ‘use’ some word of mentalese calls for brain surgery, not a mere change in the conventions of usage. There is of course an important problem of reference for the words of mentalese (if such there be), and more generally for how we refer in thought, but, as signalled earlier, this is not the problem of reference that the Lockean description theory we are defending is concerned with.

The objection that reference is in part a matter of how things are outside the head

Immediately after the passage quoted above, Devitt goes on to say

This deep failing of description theories is brought out by Hilary Putnam’s slogan... “‘Meanings just ain’t in the *head*’.... The association of descriptions with a word is an inner state of the speaker. No such inner state can make the word refer to a particular referent. For that we must look for some relation that language and mind have to things outside themselves—we must look for an external relation.^{ix}

We can say straight off that there must be something wrong with this objection. It is *too* good. The description theory is correct for some words. We do use some words simply as abbreviations for definite descriptions, and the question as to what they refer to is nothing other than the question as to what has the property associated with the abbreviated description.^x This is granted by critics of the description theory. As Saul Kripke notes, you might say to yourself ‘By “Gödel” I shall mean the man, whoever he is, who proved the incompleteness of arithmetic’, and grants that in this case ‘Gödel’ in your mouth would refer to Schmidt if it was indeed Schmidt who proved the theorem. Kripke is explicit that his claim is that, as a matter of fact, this is not what most of us do; not that we could not have done it, and not that we never do it.^{xi} Indeed, it would be extraordinary if the description theory were not true for some words. As noted earlier, we use flags to give information about how things are, and this depends on there being a conventional association between flag configurations and properties in the minds of those who use and see the flags. It would be extraordinary if we never did with words what we do do with

flags.

Moreover, the description theory *does* give a major role to the world in settling what refers to what. According to it, what settles the reference of words like ‘London’ and ‘cow’ is the combination of the properties associated with the words and a fact about the world, namely, what in the world has the properties. It is, therefore, false that the theory deprives the world of a role in settling what a term refers to.

It might be objected that I have misunderstood the sense in which Devitt is insisting that the world must play a major role in settling reference. Of course, what a word refers to *at* a world on the description theory depends on the nature of that world. The trouble is, rather, that the world does not play a role in settling the reference *conditions* of a word on the description theory in the sense of what the word refers to at w for any w . But in fact description theorists *per se* are not committed to denying the actual world a role in settling reference conditions. Description theorists should (and do) allow that some words function like rigidified definite descriptions, and the reference conditions of rigidified descriptions depend on the nature of the actual world.

The core idea behind the description theory is that a term refers to that which has the property or properties the speaker associates with the term. One way to spell this out is by saying that, at any world w , T refers to x just if x has the associated property or properties in w . Another way, equally faithful to the core idea, is by saying that, at any world w , T refers to x just if x is the thing with the associated property or properties *in the actual world*. The second way of spelling matters out is the right way if T is a rigidified definite description, a description like ‘the actually tallest person in 1991’. For the actually tallest person at any world is whoever is the tallest at the actual world. But if description theorists should (and do) count rigidified definite descriptions as referring expressions, then it is transparent that they should (and do) allow that reference conditions may depend on how things actually are. The reference of a rigidified definite description at any world w is precisely something that depends on how things actually are.

The objection that the description theory is a species of eliminativism about reference

If the language we are concerned with has predicates for the various properties the description theory holds serve to secure the reference of the names in that language, the description theory becomes in effect the view that names, be they of things or kinds, are really a style of abbreviated definite description: if the language has F as the predicate for the property that settles the reference of N , the description theory amounts to the view that N is an abbreviation of ‘the F ’. The objection is that Russell teaches us that definite descriptions are not referring expressions, that they do not refer, and so the description theory eliminates the very phenomenon it is supposed to be analysing.

I accept Russell’s account (when due account is taken of various niceties to do with conversational context) but I think that it is misleading to say that it means that definite descriptions are not referring expressions. Russell famously shows us how to give the truth conditions for certain sentences containing definite descriptions in terms that do not contain the definite descriptions themselves and do not contain anything equivalent to them as semantically significant units. According to his account,

(1) The F is G

iff

(2) There is an F which is G and every F is identical with it.

But to think that this shows that definite descriptions are not referring expressions, that they do not refer, misunderstands the sense of ‘reference’ in which the description theory is a theory of reference. Nothing in Russell’s theory goes against the fact that the words ‘the tallest person alive in 1990’ are quite distinct from the person who is the tallest person alive in 1990, and that there is some important relation between the words and the person which warrants a name. It is *this* relation that the description theorist calls ‘reference’, and it is this relation that the description theorist sees as holding also between names and the things they name. It is, of course, open to anyone to withhold the word ‘reference’ for this relation on the ground that we can give a contextual analysis of definite descriptions—or

perhaps for some other reason—but then they will need a new term for the relation between the words ‘the tallest man alive in 1990’ and the tallest man in 1990. But now the objection has come down to a claim about how best to label theories; it is no longer an objection to the description theory as such.

Moreover, an equally ‘contextual’ account in the spirit of Russell can be offered of rigidified definite descriptions, as follows.

(3) The actual F is G

is true at w iff

(4) w is the actual world and there is an F in w which is G in w and every F in w is identical with it, or w is not the actual world and there is something in w which is F in the actual world and every F in the actual world is identical with it, which is G in w .

So one who says that definite descriptions are not referring expressions on the basis of a Russell style story about them is committed to saying that rigidified definite descriptions are not referring expressions either. This is hard to believe. It means for starters that we cannot explain what a rigidified definite description is by saying that it is one which *refers* at every world to what it *refers* to at the actual world—it never refers at all, on the view in question.

It might be responded that *of course* there is *some* relation between the words ‘the tallest man’ and the tallest man, and that we are free to call this relation ‘reference’. But it differs markedly from the relation between, on the one hand, the rigidified definite description ‘the actual tallest man’ or some other descriptive name, or a proper name of the man, and, on the other hand, the man himself.^{xii} More generally, there are many differences between definite descriptions, on the one hand, and rigidified definite descriptions and names, on the other—a familiar example is their behaviour in counterfactual and modal contexts. The issue, then, the response might run, is whether

there is enough in common between the cases to justify describing them alike; whether there is, as Gareth Evans puts it, ‘any natural semantical kind’ here.^{xiii}

One reply to this response is that, whatever their differences, definite descriptions, descriptive names, and proper names are united in being markedly different from words like ‘and’, ‘is’, and ‘to’. But the really important point is that we do not have in this response a *free standing* objection to regarding definite descriptions as referring terms: it is a good objection to treating definite descriptions as referring terms only if there is *independent* good reason to reject the description theory of reference. For if the description theory of reference is correct, there is an obvious natural semantic kind in common between definite descriptions, descriptive names and proper names: they are all associated with properties in the minds of their users, and hence they have a special relation to whatever possesses the associated properties. In consequence, we cannot have in the ‘there are many differences’ point an objection to the description theory of reference. For only if the theory is false, is it at all plausible that the differences are enough to stop us acknowledging that there is an important semantic kind in common to definite descriptions, descriptive names, and names—a semantic kind that we might as well describe as their all being referring expressions.

It is often said that the description theory of reference holds that names are synonymous with definite descriptions. But in fact the key claim in the theory is that the fundamental mechanism whereby names and definite descriptions secure reference is the same, namely, via possession of associated properties—it is only in this sense that names (in rich enough languages) can be thought of as abbreviated definite descriptions. This is consistent with the referential behaviour of names and definite descriptions differing in various contexts in ways that would warrant saying that they differ in meaning. Take, for example, the widely observed differences between the behaviour of names and definite descriptions under counterfactual assumptions: the reference of a name does not change under counterfactual assumptions, the reference of a definite description may change.^{xiv} But this does not affect the basic claim that reference is via associated properties in both cases; it simply means that in the case of names, the reference under counterfactual assumptions is to what has the associated property *in fact*, whereas the reference of a

definite description may be to what has the property under the counterfactual assumption. Once we see that the essential claim of the description theory of reference is not a synonymy one, but instead about the sharing of a fundamental mechanism, it is clear that the manifest variety among referential devices concerning their behaviour in various contexts is no reason to deny the underlying semantic unity postulated by the description theory.

The objection from ignorance and error^{xv}

Surely, runs this objection, there are cases where it is intuitively clear that speakers refer to something *O* by their use of *T*, where one or both of the following applies: (a) a speaker knows nothing, and maybe knows that they know nothing, that individuates *O* from many other things, and (b) most of what a speaker believes about *O* is wrong. But if reference is by possession of associated properties, then (a) these properties must be sufficient to individuate *O*—the user of *T* must not be ignorant of what individuates *O*, and (b) *O* must have these properties—the user of *T* cannot be in error about *O* having them.

The cases offered to support the claim about ignorance of individuating properties all seem to me to overlook obvious candidates to be the needed individuating properties. Hilary Putnam claims that he does not know what separates beeches from elms but insists that he succeeds in referring to beeches when he says, say, that he does not know how beeches differ from elms.^{xvi} I agree that he does refer to beeches, but point out that he does know how they differ from elms: only they are called ‘beeches’ by the experts in his language community. Putnam responds that, because the word in French for beech is different from our word, this reply would commit description theorists to holding that a not very knowledgeable, monolingual English speaker’s concept of a beech will be different from that of a not very knowledgeable, monolingual French speaker. But how is this a problem? Peoples’ concepts of one and the same thing can and do differ, and it is hard to see why this should not count as a case.

There are though cases where appeal to knowledge that certain words are used by experts does not turn the trick for the description theory. Perhaps I read in the paper about

someone called ‘Smith’ who was robbed last night in Washington. Can’t I refer to him even though there may well be more than one person called ‘Smith’ robbed in Washington last night? And there are examples of meeting someone late on at a party. It seems that I can refer to him next morning—perhaps in some such sentence as ‘I wonder if I insulted that man I only vaguely remember who I met as I was finishing the bottle’—despite his having no obvious individuating marks that I can recall. But in fact I do know something that individuates what I refer to in cases of these kinds. They are the causal source of the information-carrying trace that I am presented with. Out of the possibly many Smiths robbed in Washington last night, there will be one that is the causal origin of the story in the paper, in the sense that it is his being robbed that lead to the passage of prose in the paper in front of me. Or, rather, there had better be if the case is to be a test case for theories of reference: if it turns out that the reporter muddled up a number robbed Smiths, and the passage of prose can equally be thought of as sourced in any, or none, of them, we lose the intuition that there is one particular Smith that I refer to. Similarly, there will be only one person who is the right kind of causal origin of my vague, somewhat disturbing memory of the party that is prompting my morning after reflections (and who will, most likely, also be the only person I would recognise as such were I to meet him). Hence, the description theorist can explain how I manage to refer in this kind of case by appealing to an association with the property of being a certain kind of causal origin (and maybe the property of being apt to be recognised as such by me).

Suggestions of these kinds in support of a description theory of reference are not new.^{xvii} Their impact has, however, been lessened by the belief that they face serious problems. I will consider three of the problems most often brought up.

First, the appeal to language use by experts suggests an implicit circularity. It looks circular to say that ‘beech’ in the mouths of the ignorant refers to whatever ‘beech’ in the mouths of the experts *refers* to, and isn’t that what the ‘reference borrowing’ suggestion really amounts to?^{xviii} However, the circularity is not vicious. We can spell the suggestion out without explicit mention of reference. The suggestion is simply that the property the ignorant associate with the word ‘beech’ is having the property, whatever it is, that the experts associate with the word ‘beech’. Hence, on the description theory, ‘beech’ in their

mouths refers to whatever has the property the experts associate with the word ‘beech’—which ensures, as it should, that ‘beech’ in the mouths of the ignorant has the same reference as it has in the mouths of the experts.

Secondly, it is observed that it may be unclear who the experts are in the sense that it is unknown to the ignorant users, and yet reference still occurs.^{xix} Many folk refer to quarks when they use the word ‘quark’, despite the fact that they do not know whether it is a term from physics or from biology, and so do not know which department contains the experts. For when such a person asks, ‘Should I go to the physics or biology department to find out about quarks?’, it is plausible that they are asking a question about *quarks*. However, these folk do know that there are some experts somewhere or other, and that these experts lie at one end of a reference borrowing chain that has whoever they themselves borrowed the term from at the other—or at least this had better be the case, for otherwise the example is no longer one where it is plausible to say that these folk refer to quarks. But now we can specify the property these folk associate with the word ‘quark’. It is having the property the group of users of the word ‘quark’ that they are borrowing from associate with the word ‘quark’. And what property does this second group of users associate with the term ‘quark’? Either they are the experts, in which case it is property *Q* (whatever that is, I am not one of the experts) or they are not the experts, in which case it is having the property some third group of users associates with ‘quark’. And what property does this third group of users associate with the word ‘quark’? Either they are the experts, in which case it is property *Q*, or they are not the experts, in which case it is having the property some fourth group of users associate with the word ‘quark’.... We have, that is, a recursive story akin to the familiar recursive account of a wff in logic.

The third problem raised turns on the point that the key issue is not whether there is some individuating property of the thing referred to. It is whether there is some individuating property of the thing referred to which is associated with the word in question by the speakers (and maybe hearers) in question. The objection is that the kinds of properties we have mentioned are fancy ones that philosophers of language and unusually alert members of the folk might think of; they are not properties ordinary, folk speakers associate with words like ‘beech’ and ‘quark’. A quick reply to this objection is

that the folk often say things which make it clear that they are well aware of these properties and which strongly suggest that they are relying on them to secure reference. People who do not know much physics, and know that they do not know much physics, often ask questions like, Is it established for sure that quarks exist? When asked precisely what question they are asking, they answer that they are asking about the things physicists use the word ‘quark’ for. Or think of what has happened with the spread of computer speak. It is a commonplace that people say things like ‘I haven’t a clue what RAM is, but I know that the new PC I am buying has 32 of whatever it is that computer people use the term for’.

However, we need to say more by way of reply. Although the folk are aware of the relevant individuating properties and say things that suggest they are relying on them to secure reference, they are not much good at articulating them in detail (nor, if it comes to that, are defenders of the description theory of reference—we do a fair bit of hand-waving). The longer reply is that this failure to articulate the relevant property or properties in detail is no objection to the description theory *provided* that what is meant by the expression ‘properties associated with a word or phrase’ in statements of the description theory is understood in the right way.

Sometimes it is obvious which properties are associated with a word. Perhaps the speaker tells us loud and clear. But typically the association is implicit or tacit rather than explicit. It is something we can extract in principle from speakers’ patterns of word usage, not something actually explicitly before their mind when they use the words. I know this way of putting things—familiar though it is—will ring some alarm bells. Some will want to say that if the association is in the mind, as the description theory says, it must be explicit. They think of appeal to the implicit or tacit in this context as a kind of cheat—a way of saying something and then taking it back.

However, there is a way of being implicit and yet before the mind in the relevant sense that is no great mystery. Consider the situation good logic students find themselves in before they are given the recursive definition of a wff. They cannot specify what it is to be a wff, but they can reliably classify formulae into wffs and non-wffs. But it would be a mistake to see their ability as like that of chicken sexers. Chicken sexers have no idea

which property triggers their reliable classifications (although they know, of course, that the property is correlated with the sex of the chicken). By contrast, logic students can say for any ill-formed formula what triggers their judgement that it is ill-formed. When presented with $(p \vee q)$, they do not say that they can see that it is ill-formed but cannot say where the problem is. They know exactly where the problem is and how to fix it—add a RH bracket after the ‘q’. Similarly, they know what changes to a particular wff would make it ill-formed. They are in the following position: for each particular example, they can say whether or not it is a wff, and why; but they cannot give in words a story that covers all cases. The same is true for all of us in our judgements of grammaticality. We can say, for particular examples, whether and why they are or are not grammatical, but cannot give the general story in words. In the case of wffs, many of us know what the general story is, and tell students when we give the recursive definition of wff-ness; in the case of grammar, we (with the possible exception of Quirk and Greenbaum) don’t know the general story.

Description theorists can and should say essentially the same about the sense in which speakers associate properties with words. If you say enough about any particular possible world, speakers can say what, if anything, words like ‘water’, ‘London’, ‘quark’, and so on refer to in that possible world. (This does not mean that there is always a definite answer: sometimes saying what ‘London’ refers to in a certain possible world will amount to saying that it is indeterminate what if anything it refers to in this world.) Our ability to answer questions about what various words refer to in various possible worlds, it should be emphasised, is common ground with critics of the description theory. The critics’ writings are full of descriptions (*descriptions*) of possible worlds and claims about what refers, or fails to refer, to what in these possible worlds. Indeed, their impact has derived precisely from the intuitive plausibility of many of their claims about what refers, or fails to refer, to what in various possible worlds. But if speakers can say what refers to what when various possible worlds are described to them, description theorists can identify the property associated in their minds with, for example, the word ‘water’: it is the disjunction of the properties that guide the speakers in each particular possible world when they say which stuff, if any, in each world counts as water. This disjunction is in their

minds in the sense that they can deliver the answer for each possible world when it is described in sufficient detail, but is implicit in the sense that the pattern that brings the various disjuncts together as part of the, possibly highly complex, disjunction may be one they cannot state. This is not to say that, after reflection on their classifications in the various possible cases, perhaps aided by doing a course in the philosophy of language, they won't be able to make good stab at stating the pattern; something like: belonging to the kind which most of the clear, potable samples, acquaintance with which lead to the introduction of the word 'water' in our language, is roughly right—and if you describe a case that this formula fails to cover, you do not show that there is no pattern, but that my stab at it was not good enough.^{xx}

We can now deal quickly with the error side of the objection. The claim is that speakers may have most of the properties of *O* wrong and yet still refer to *O* by their use of *T*. But what matters for successful reference is that *O* has the properties speakers *associate* with *T*, and this is consistent with *O*'s lacking most of the properties speakers think it has. There may be only one property speakers associate with *T*, in which case they refer to *O* when they use *T* provided just that *O* has that property; it will not matter if everything else they believe about *O* is mistaken.

The Twin Earth objection

There are two different ways of thinking of Twin Earth. On one, it is part of our (the actual) possible world; on the other, it is part of another possible world. The two ways raise quite different issues for the description theory.

Suppose that Twin Earth is on the other side of our sun and has, as is traditional, a population that speaks something very like English, and stuff superficially very like our water which plays on Twin Earth a role very like the role water plays on Earth. This stuff is called 'water' by Twin Earthians but has a quite different chemical composition from our water: it is XYZ, not H₂O. One Twin Earth objection is that 'water' in our mouths does not refer to XYZ despite the fact that XYZ has so many of the superficial and functional properties of water.

This is only an objection to the description theory if the description theory is committed to saying that ‘water’ in our mouths refers to XYZ. But, as we have just been emphasising, the description theory is committed to ‘water’ in our mouths referring to whatever has the property we associate with ‘water’, and the test for being the property we associate with ‘water’ is that it is the, possibly disjunctive, feature common to the possible cases we describe as water. The very fact that we resist saying that the water-like stuff on Twin Earth is water shows that it does not have the property we associate with ‘water’. There is, as we noted a moment ago, no way that an appeal to intuitions about possible cases can refute the description theory; at most it can tell us that we, or someone, got the associated property or properties for some word wrong. In fact, what we learn (and this is an important thing to learn; I am not decrying the interest of Twin Earth) is that acquaintance is an important member of the properties we associate with ‘water’ (and lots of other words). The reason XYZ is not what *we* refer to when *we* use the word ‘water’ is that it is not the water-like kind *we* are acquainted with. We have never ourselves come across XYZ. But if we had, if some of the water-like stuff around us was H₂O, and some XYZ, and there was no reason to think of one as ‘fools’ water, then we would have been in the same situation with respect to ‘water’ as we in fact are with the word ‘jade’—two different kinds would have been covered by the one word.

The other Twin Earth objection invites us to consider the possible world where XYZ is like H₂O in both superficial qualities and role, *and* is the stuff we are acquainted with in this non-actual world. In this version, Twin Earth is not a place remote from Earth in the actual world but is in effect Earth, our Earth, in a possible world where our Earth has XYZ as the stuff that is the clear odourless, potable liquid that we are acquainted with and call ‘water’. The objection now is that ‘water’ in our mouths does not refer to XYZ in this world, and yet XYZ has in this world any and every property that might plausibly be held to be associated with ‘water’ and, in particular, has the property of being the water-like kind we are acquainted with in this world.

The reply to this objection is that the moral is as before: we are learning something about the relevant associated properties from these reflections on how to describe possible cases, not that reference does not go by associated properties. Twin Earth, in this second

version of the objection, is given to us as a *counterfactual* world, a world other than our actual one; and what we learn from the fact that ‘water’ does not refer to XYZ in this counterfactual world is that the property associated with ‘water’ is the property of being the kind that *actually* is the water-like kind of our acquaintance. We learn that what matters is not the property XYZ has in the counterfactual world, but the property it has, or rather fails to have, in the actual world. It is because XYZ is not the water-like kind of our acquaintance in the actual world that the word ‘water’ does not refer to it. As you would expect if this account is correct, if Twin Earth is thought of as how our world actually is, instead of as a counterfactual world, ‘water’ does refer to XYZ in it. Perhaps we are contemplating the possibility that some terrible mistake was made by the founders of modern chemistry, or that some fabulous fraud was perpetrated—they knew all along that it was XYZ but decided to falsify the first text books and rely on lazy copyings from them to later ones to perpetuate the fraud—then what else is there to say but something like, “If *that’s* how things are, hard though it is to believe that the error or fraud could have remained undetected for so long, then water is XYZ”? And, in this case, it is of course XYZ, not H₂O, that has the property of being the actual water-like kind of our acquaintance.

The objection from divergent sets of associated properties

I know of no statement of this objection in the terms that follow, but it is suggested by §2.3 of *The Varieties of Reference* and it comes up regularly in discussion. The objection takes off from the point that very often the properties associated with a given word by different people differ markedly: names for towns and people are particularly clear cases. What then are the associated properties that, on the description theory, settle the reference of such words?

This is only an objection if there is a problem about holding that, in such cases, the answer to our rhetorical question varies from speaker to speaker. It sometimes seems to be thought that there is a serious problem with allowing that different, competent speakers of the same public language might associate with a proper name like ‘Paris’ very different

sets of properties. But it is hard to see what this problem might be. After all, by suitable mutual quizzing we could identify these differences, if we thought it worth the trouble. It would, of course, be a nuisance if the city that had the properties I associate with ‘Paris’ was a different city from the one that has the properties you associate with the word ‘Paris’—arrangements for us to meet expressed using the word ‘Paris’ would go badly astray. But this is unlikely to be the case, and, moreover, we know that this is unlikely to be the case.

Communicating how things are by the use of words turns on associations between words and properties in the minds of speakers and hearers, as we said near the beginning. In the ideal case, the properties associated by speaker and hearer will be the same, but in practice this is often far from the case. As long as we can bring the properties associated by speakers and hearers into line when it matters, no great harm results from divergences from the ideal, especially when the focus is more on getting our paths to cross than on the nature of the place where our paths cross.^{xxi}

Sometimes the objection from divergent sets of properties is phrased on the assumption that the set of properties associated with a name is the meaning of that name according to the description theory, and the objection is then put in some such terms as, “But, by definition, persons who share a language give words the same meaning. So the description theory is refuted by the fact that people who share a language do sometimes associate quite different properties with a word in their common language.” But if that is the criterion for sharing a language, we should say that few people share a language, though very many go close enough for it not to matter that they don’t.

The objection from the history of science

This objection is that the history of science provides us with many cases where the description theory gets the reference of various terms wrong. There are, it is argued, many cases where advances in scientific understanding lead us to revise our views concerning the reference of some term in a way that cannot be explained in terms of the description theory. An example is the term ‘liquid’. The property we folk associate with the term

'liquid' is flowing. This would suggest that we could say *a priori* that glass is not a liquid (at room temperature) simply on the basis that it does not flow (at room temperature), but in fact science tells us that glass is a liquid.

The quick reply to this objection is that glass *is* a solid and not a liquid, on the folk meanings of the terms. There is, however, more to say. The first time it is explained to us in the science classroom that glass is a liquid, our response is *not* that this is a piece of re-definition, justified by the fact that the new definition picks out a more theoretically interesting class; it is that glass's being a liquid is an interesting discovery of science. (Or anyway this was my reaction and, as far as I can recall, that of the others in my class.) How come, if, as I insist, glass is not a liquid on the folk meaning? The answer is that context affects the meanings we give terms, as has been widely noted. What counts as a square in geometry is very different from what counts as a square in day to day talk. In particular, the very fact that one is in the science classroom affects the properties associated with terms like 'liquid' and 'solid'. The term 'solid' becomes associated, by the very act of entering the classroom, with having the scientifically interesting property that typically explains resistance to deformation and failure to flow. What you then learn in class is that this property is that of having a certain kind of crystalline structure, and that glass and tar lack this structure. They have the scientifically interesting property that typically explains flowing, and they therefore count as (super-cooled) liquids in the classroom. Similarly, a discussion *with a biologist* of whether a whale is a fish automatically becomes a discussion of whether a whale has the biologically important properties typical of things that live in the sea. This is the property we folk associate with the term 'fish' in this kind of context; the role of biologists, once they have set the context, is then to tell us that, as it happens, although whales typically live in the sea, they do not have the biologically important property of things that typically live in the sea, and so are not fish.^{xxii}

A similar point applies to an example like 'kinetic energy'. It is sometimes suggested that in Newtonian physics the property associated with the term is $1/2mv^2$, and so that the description theory cannot explain how Einstein managed to show that kinetic energy is not $1/2mv^2$.^{xxiii} However, when you read Newtonian physicists it is clear that

their equation of kinetic energy with $1/2mv^2$ was the outcome of sophisticated theory. It was not a stipulation in the sense that the metre bar in Paris was. The property they associated with the term ‘kinetic energy’ was that of playing a certain theoretical role, a role which they were convinced was played by $1/2mv^2$. What happened when relativity theory came along is that this role was shown to be occupied by something other than $1/2mv^2$, though it approximates it closely when v is not large; or, better, that nothing occupies exactly the role Newtonians gave $1/2mv^2$, but the relativistic expression occupies a role close enough for us to say that it is kinetic energy. Of course, it is open to any Newtonian physicist to insist that what *they* meant by ‘kinetic energy’ was $1/2mv^2$, and that in consequence what *they* said when Einstein came along was not that kinetic energy had been shown to be something other than $1/2mv^2$, but rather that Einstein had shown that kinetic energy did not play the theoretical role they thought it did. I think these remarks would reveal their usage to be non-standard, but so what?

A blind alley

I have sometimes met the following objection. “It is obvious that we use language to convey information about how things are, and that requires associations between words and properties. But we also use language to convey information about *objects*, and that requires associations between objects and words, rather than between properties and words. The Lockean version of the description theory that you are defending has a gaping hole in it—a hole which has to be repaired with an account of reference that does not appeal to associated properties.”

My reply is that you cannot give information about objects without giving information about their properties. Finding, identifying, or locating an object, and, more generally, making an object a subject of discussion, is necessarily finding, identifying, locating, or making it a subject of discussion *under one or more of its guises*. This is not to advance the controversial view that objects are bundles of properties; it is to insist that we access objects via their properties.

There are objections to the description theory of reference which I have failed to address, and no doubt advocates of the objections I have discussed will have objections to my objections to their objections. But I hope that I have done enough to make some re-think their view that the description theory of reference belongs with the great dead theories of the past.

Frank Jackson

Philosophy Program

Research School of Social Sciences

Australian National University

ACT 0200

- Braddon-Mitchell, David and Frank Jackson (1996), *The Philosophy of Mind and Cognition*, Oxford, Basil Blackwell.
- Devitt, Michael (1996), *Coming to our Senses*, Cambridge, Cambridge University Press.
- Devitt, Michael and Kim Sterelny (1987), *Language and Reality*, Oxford, Basil Blackwell.
- Evans, Gareth (1982), *The Varieties of Reference*, Oxford, Clarendon Press.
- Kripke, Saul (1980), *Naming and Necessity*, Oxford, Basil Blackwell.
- Kroon, Fred (1978), 'Causal Descriptivism', *Australasian Journal of Philosophy*, 65: 1–17.
- Lewis, David (1969), *Convention*, Cambridge, Mass., Harvard University Press.
- Lewis, David (1979), 'Score-Keeping in a Language Game', *Journal of Philosophical Logic*, 8: 339–59.
- Lewis, David (1984), 'Putnam's Paradox', *Australasian Journal of Philosophy*, 62: 221–36.
- Linsky, Leonard (1977), *Names and Descriptions*, Chicago, Chicago University Press.
- Locke, John (1690), *An Essay Concerning Human Understanding*.
- Neale, Stephen (1990), *Descriptions*, Cambridge, Mass., MIT Press.
- Putnam, Hilary (1962), 'The Analytic and the Synthetic', in Herbert Feigl and Grover Maxwell, eds, *Minnesota Studies in the Philosophy of Science*, III, Minneapolis, University of Minnesota Press.
- Putnam, Hilary (1975), 'The meaning of "Meaning"', in *Mind, Language and Reality*, Cambridge, Cambridge University Press.
- Searle, John (1983), *Intentionality*, Cambridge, Cambridge University Press.
- Slote, Michael (1966), 'The Theory of Important Criteria', *Journal of Philosophy*, 63: 211–24.
- Strawson, P. F. (1959), *Individuals*, London, Methuen.
- Strawson, P. F. (1971), *Logico-Linguistic Papers*, London, Methuen.

ⁱ I am much indebted to Sam Guttenplan, Richard Holton, and especially David Braddon-Mitchell and David Lewis.

ⁱⁱ See, e.g., Locke (1690), book III, Kroon (1978), Linsky (1977), Searle (1983), Lewis (1984), Strawson (1971), Braddon-Mitchell and Jackson (1996), Neale (1990).

ⁱⁱⁱ Locke (1690), book III, ch. 2 sometimes puts matters in a way that invites the first reading.

^{iv} I am indebted here to Jakob Hohwy.

^v Some description theorists hold that adjectives like ‘square’ refer to the property associated with them; some that they refer to the things that have the property; some that they refer to the set of things that have the property; and some that there is no substantive issue here: we will confine our attention to nouns and noun phrases.

^{vi} Devitt (1996), p. 159. See also Devitt and Sterelny (1987).

^{vii} Locke (1690), book III, ch. 2, § 2. I am indebted here to discussions with Monima Chadha.

^{viii} For how a convention might be implicit, see Lewis (1969).

^{ix} Devitt (1996), p. 160. Incidentally, I am not sure why Devitt says that the ‘association of descriptions with a word is an inner state of the speaker’—I happen to agree but am surprised that the point should be taken as obvious given the popularity of externalist accounts of psychological states.

^x If you hold that definite descriptions do not refer, the point could be made with descriptive names or rigidified definite descriptions. I argue against the view that definite descriptions do not refer below.

^{xi} Kripke (1980), p. 91

^{xii} I take descriptive names to be abbreviated rigidified definite descriptions, and so will, from now on, treat them together. As far as I can see, nothing important for what follows turns on this.

^{xiii} Evans (1982), p. 57.

^{xiv} Evans (1982), p. 57.

^{xv} This is perhaps the most common objection to the description theory. I take the name of the objection from Devitt and Sterelny (1987).

^{xvi} Putnam (1975), p. 226..

^{xvii} The earliest stressing of the importance of causal links I know is Strawson (1959), ch. 1.

^{xviii} This issue is discussed at length by Kripke (1980), lecture II. See also Devitt and Sterelny (1987), p. 50.

^{xix} See, e.g., Devitt and Sterelny (1987), p. 50.

^{xx} Or, maybe, that the way you introduced the story ‘disturbed’ the pattern. Description theorists can and should allow that the associations between words and properties are not set in stone.

^{xxi} When the properties diverge, description theorists can distinguish speaker reference from hearer reference in the obvious way. The rough definition we gave near the beginning was, of course, of speaker reference.

^{xxii} The example is from Slote (1966). The role of context is discussed in detail in Lewis (1979).

^{xxiii} The example is often sourced to Putnam (1962) but it should be noted that he is using it to slightly different effect.