

ON RUSSELL'S THEORY OF DEFINITE DESCRIPTIONS

Wylie Breckenridge

This essay is about Russell's theory of definite descriptions and how it deals with those that denote nothing. Russell thought that most proper names are abbreviated definite descriptions and so lie within the scope of the theory as well. Although that's an interesting aspect, I won't be considering it here. The essay has three parts. In the first I describe Russell's motivation for developing the theory. In the second I describe the theory and how he justifies it. In the third I look at some objections that have been raised against it and offer some responses on Russell's behalf.

PART I: RUSSELL'S MOTIVATION

Some of our statements contain definite descriptions: 'The queen of England loves croquet' contains the definite description 'The queen of England'; 'The greatest prime number does not exist' contains the definite description 'The greatest prime number'. I will not try to define what I mean by a definite description, apart from saying that I mean phrases of the form 'The so-and-so' like those above, but *not* like 'The good old days' (because those days were not old and some at least were not good) and 'The big pineapple' (because although it is big it is not actually a pineapple). If there is exactly one object that is so-and-so then I will say that the definite description *denotes* that object, or that the object is the definite description's *denotation*, or that the object is *denoted by* the definite description. If there is no object that is so-and-so, or more than one, then I will say that the definite description denotes nothing. For example, 'The natural satellite of the Earth' denotes the Moon, whereas 'The natural satellite of Mercury' and 'The natural satellite of Jupiter' denote nothing.

Russell's theory can be seen as an answer to this question: What role do definite descriptions play? Or, what are they used to do? Or, what is their purpose? The aim of this first section is to show that the question has no easy answer, and to look at what motivated Russell to come up with a theory that seems, at least initially, complicated and counter-intuitive.

An obvious answer is that the role of a definite description is purely to refer to the object that it denotes - that its task is nothing more than to pick out an object about which we wish to say something. In 'The queen of England loves croquet', for example, the role of 'The queen of England' is to pick out a person whom we wish to say stands in the relation of loving to croquet.¹ This answer seems to be wrong for at least two reasons.

First, if the task of a definite description is just to refer, then any definite description that denotes the same thing could equally well perform the task - that is, could equally well be used in place of the first. For example, 'The queen of England loves croquet' could be expressed just as well by 'The mother of Prince Charles loves croquet'. While it may be so in this case, there are others in which it is not. The truth of 'The evening star

¹ I am taking referring to be distinct from denoting, although the difference is subtle and probably not important. The difference is basically this: denoting is something that a definite description might do, whereas referring is something that a definite description might be *used* to do.

is the morning star' was an interesting and surprising astronomical discovery. But 'The evening star' denotes the same thing as 'The morning star' (the planet Venus), so if their role is purely referential then it shouldn't matter if we express this instead as 'The evening star is the evening star'. But it does - this last statement can be known to be true without the help of astronomy. So there is something *other* than the denotation of a definite description that has *some* effect on its containing statement. It seems that definite descriptions are sometimes used to do more than just refer.

Second, suppose someone asks me whether 'The queen of England loves croquet' is true. I understand the statement and what I have to do in order to find out its truth value - ask the Queen, or ask someone who knows the Queen, or look it up in a book, or something like that. So it seems that the definite description has performed its task here. But suppose that there is no queen of England - that her purported existence has been a massive hoax. Then the 'The queen of England' cannot refer to the object that it denotes because there is no such object - it denotes nothing. But because it performed its task, that task cannot have been to refer. It might be objected that if the 'The queen of England' does not refer then no truth value can be assigned to the statement, and so the definite description did not perform its task after all. If so, then consider instead 'The greatest prime number does not exist'. Mathematicians believe that this is true *because* they believe that 'The greatest prime number' denotes nothing. Not only can they understand the statement, but they can give it a truth value, so surely the definite description has performed its task here. And yet it cannot refer. So it seems that at least sometimes it is not the task of a definite description to refer.

As a solution to this second problem, Meinong (1904) proposed that definite descriptions cannot fail to refer because they *always* denote an object. His idea was that the objects denoted by 'The King of England', 'The greatest prime number', and 'The round square cupola on Berkeley College' *merely* exist, whereas those denoted by 'The queen of England' and 'The least prime number' exist but also *subsist*. So existence, Meinong suggested, comes in two varieties. Russell rejected this idea, for two reasons. The first was that holding it could be seen as a "failure of that feeling for reality which ought to be preserved even in the most abstract studies" (1919, p. 169). He urged that "there is only one world, the 'real' world" (loc. cit.), and that "a robust sense of reality is very necessary in framing a correct analysis of propositions about unicorns, golden mountains, round squares, and other such pseudo-objects" (op. cit. p. 170). The second was that some of the supposedly existent objects defy the law of non-contradiction: the round square is round and also not round, for example. He says that this is "intolerable" and that "if any theory can be found to avoid this result, it is surely to be preferred" (1905, p. 45).

In response to the first problem, Frege (1892) offered quite a different theory. He suggested that definite descriptions (and other linguistic items as well) have a 'sense' (something like what we might call 'meaning') and a 'reference' (what we might call 'denotation'). A grammatically well-formed definite description always has a sense, and sometimes has a reference. The sense can be thought of as a way of picking out an object which may or may not succeed. If it does, then the object is its reference. If it doesn't, then it has no reference. For example, 'The queen of England' is a way of picking out an object that *does* happen to succeed, and the object it picks out is its reference. On the other hand, 'The King of England' is a way of picking out an object

that *doesn't* happen to succeed, and it has no reference. But both have a sense because they are grammatically well-formed.

Armed with this notion of sense, it is open to Frege to say that the task of a definite description is not to refer but to provide a sense, which may or may not come with a reference. He thus avoids the second problem - 'The queen of England' has a sense whether or not it has a reference, and so it can perform its task in 'The queen of England loves croquet' whether or not there actually is a queen of England. And he also avoids the first problem - although they have the same reference, 'The evening star' and 'The morning star' have different senses, and that is what accounts for the cognitive difference (the different degree of interest they carry) between 'The evening star is the morning star' and 'The evening star is the evening star'.

But despite its intuitive appeal, Russell rejected Frege's theory. In 'On Denoting' (1905, pp. 48-51), he argues that the relation of the meaning (sense) to the denotation (reference) of a definite description involves difficulties which are "sufficient to prove that the theory which leads to such difficulties must be wrong". His arguments are very difficult to follow, but one of them goes like this. Suppose C is a definite description (Russell talks more generally about 'denoting complexes', but I'll stick to the special case of definite descriptions). According to Frege, C is used to talk about the denotation of C, and 'C' is used to talk about the meaning of C. That is, 'C' denotes the meaning of C. But, Russell says, 'C' also denotes C. So the meaning of C is just C. There cannot be three different entities - C, the meaning of C, and the denotation of C. There is just C and its denotation. Unfortunately I cannot understand Russell's other arguments, but his conclusion is quite clear: he says that "This is an inextricable tangle, and seems to prove that the whole distinction of meaning and denotation has been wrongly conceived" (p. 50), and that "the point of view in question must be abandoned" (p. 51).

So what is Russell's theory?

PART II: RUSSELL'S THEORY

Russell's theory (1905 for original presentation, 1919 for a clearer account) is quite different to that in which the role of a definite description is purely to refer, and to Frege's, in which even though it might provide a reference its role is to provide a sense. According to Russell, definite descriptions are *never* used to refer and have no sense or meaning of their own. He says (1905, p. 51) that although 'Scott was a man' and 'The author of *Waverley* was a man' have the same grammatical form they have quite different logical forms. 'Scott was a man' is a statement of the logical form 'x was a man' and has 'Scott' for its logical subject, but 'the author of *Waverley* was a man' is *not* a statement of the logical form 'x was a man' and does *not* have 'the author of *Waverley*' for its logical subject.² If we have thought so in the past then it's because we have been "mislead by grammar" and have "regarded grammatical form as a surer guide in analysis than, in fact, it is" (1919, p. 168). He says that definite descriptions are really *quantifier* phrases that belong essentially to whole statements. Russell's central claim is that a definite description 'the Φ ' (for some predicate Φ) is to be interpreted not as

² It's not clear that Russell should describe 'Scott was a man' and 'The author of *Waverley* was a man' as having different logical forms, given that he thought that 'Scott' is an abbreviated definite description. But we can ignore that issue here - nothing important hangs on the contrast.

referring (or attempting to refer) to the one and only Φ , but (in part) as asserting the existence of one and only one Φ . The precise interpretation depends on the form of the statement in which it occurs. Russell distinguishes three forms (Φ and Ψ are predicates): 'The Φ is Ψ ', 'The Φ does not exist', and 'The Φ is not Ψ '.

Statements of the form 'The Φ is Ψ ' are to be interpreted as 'At least one thing is Φ , at most one thing is Φ , and whatever is Φ is Ψ '. Or more simply (and almost but not quite equivalently³) as 'Exactly one thing is Φ and it is Ψ '. For example, 'The queen of England loves croquet' is to be interpreted as 'Exactly one thing is queen of England and it loves croquet'.

Statements of the form 'The Φ does not exist' are to be interpreted as 'It is not the case that exactly one thing is Φ '. If Russell were to treat 'exists' as a predicate, then he would have to interpret these as 'Exactly one thing is Φ and it does not exist'. Unless there is a distinction between 'there is' and 'there exists', which Russell rejects, then this is a contradiction. But he doesn't want it to come out as a contradiction because in some cases it is believed to be uncontroversially true (as 'The greatest prime number does not exist' is believed by mathematicians). So Russell has to deny that 'exists' is a predicate, and interpret them in the way described. (He might say that these statements are another case in which grammatical form is misleading.) Similarly, 'The Φ exists' is to be interpreted simply as 'There is exactly one Φ '. So 'The greatest prime number does not exist' is to be interpreted as 'It is not the case that there is exactly one greatest prime number', and 'The queen of England exists' is to be interpreted as 'There is exactly one queen of England'.

Statements of the form 'The Φ is not Ψ ' are, he says, ambiguous. They can be interpreted either as 'There is exactly one Φ and it is not Ψ ', in which case 'The Φ ' is taken to have what he calls *primary* occurrence (we might say that 'not' has narrow scope), or as 'It is not the case that (there is exactly one Φ and it is Ψ)', in which case 'The Φ ' is taken to have *secondary* occurrence (we might say that 'not' has wide scope). The significance of this ambiguity will become apparent below. Just note here that only *one* of these interpretations is the negation of 'The Φ is Ψ ' - it's the second one, in which 'The Φ ' is taken to have secondary occurrence. The other is not the negation of 'The Φ is Ψ ' - it is contrary to but does not contradict it.

The claim that definite descriptions are never used to refer and do not have meaning of their own seems counter-intuitive. How does Russell justify it? He does so in three ways (1905).

First, he raises some problems faced by other accounts of definite descriptions. We saw those problems in part I.

Second, he appeals to the similarity of definite descriptions with indefinite descriptions - phrases like 'a man', 'some man', 'all men', etc. It is intuitively acceptable to say that in 'Some man is my father' the indefinite description 'Some man' does not purport to refer to any particular thing, and that the statement can be interpreted as 'There is least one

³ The problem with the simpler version is the word 'it'. What is it? The thing that is Φ of course. But then we have a definite description again. It's a technical problem that I think we can safely ignore here.

man that is my father'. Russell claims that '*The* man is my father' is just like '*Some* man is my father', except that it also asserts uniqueness. So it should be interpreted in a similar vein as 'There is at least one man that is my father, and there is at most one man that is my father', or as 'There is exactly one man that is my father'. In general, he claims that it is natural to move from interpreting 'Some Φ is Ψ ' as 'At least one Φ is Ψ ' to interpreting 'The Φ is Ψ ' as 'At least one thing is Φ , at most one thing is Φ , and whatever is Φ is Ψ '.

Third, he shows how his theory can solve three 'puzzles' about definite descriptions. The first is the problem about 'The evening star' and 'The morning star'. (The example that Russell uses is actually about 'Scott' and 'The author of Waverley', but I'll stick to the morning and evening stars.) The problem is that the truth of 'The evening star is the morning star' is interesting, and yet when we replace 'the morning star' by 'the evening star' (which denotes the same thing) we get the uninterestingly true statement 'The evening star is the evening star'. Russell's solution is that the apparently co-referring definite descriptions do not refer at all. The original statement is not *about* a thing called 'the morning star', it just includes a claim about the unique existence of a thing with certain properties. So we cannot make the substitution in the way suggested. The second puzzle is that some statements involving definite descriptions seem to defy the law of the excluded middle. According to it, the King of England is either bald or not bald and so at least one of 'The King of England is bald' and 'The King of England is not bald' must be true. But if we listed all of the things which are bald and all of the things which are not bald we would not find the King of England on either list (because there is no King of England). So it seems that neither is true. Russell's solution is to point out that the law of the excluded middle says that the King of England is either on the first list or not on the first list. But not being on the first list is *not* the same as being on the second list - this is the important distinction between the two interpretations of 'The King of England is not bald' that we noted above. For the law of the excluded middle to hold, it only has to be the case that the King of England is either on the first list or not on the first list (and that *is* the case). It does *not* have to be the case that the King of England is either on the first list or on the second list (just as well - because this is *not* the case). The third puzzle came up in part I as well - how can we talk about things that do not exist in order to (truthfully) deny their existence? If we can talk about them then mustn't they, in some sense, exist? (Meinong thought yes.) Russell thinks no. His solution we have already seen - to deny that in the statement 'The greatest prime number does not exist' the word 'exists' is used as a predicate. Rather, the statement should be interpreted as saying 'It is not the case that there is exactly one greatest prime number'.

That is Russell's theory and his justification for it. There are objections.

PART III: OBJECTIONS AND REPLIES

One way to object is to criticise Russell's criticisms of Meinong and Frege, thereby criticising his motivation for developing the theory. It could be said, for example, that to accuse Meinong of failing to keep a 'robust sense of reality' is just to accuse him of not sharing an ontological prejudice, and that the round square's defiance of the law of non-contradiction could just as well be taken as proof that it has the lesser of two kinds of existence, rather than as proof that the notion of two kinds of existence is absurd. But this way of objecting is not my main concern.

Another way to object is to criticize the way Russell justifies the theory. It could be said, for example, that the similarity between definite and indefinite descriptions was drawn too quickly. To recall, he claims that the former are just like the latter but with an added claim of uniqueness, so that it is natural to move from interpreting 'Some Φ is Ψ ' as 'At least one Φ is Ψ ' to interpreting 'The Φ is Ψ ' as 'At least one thing is Φ , at most one thing is Φ , and whatever is Φ is Ψ '. It seems, rather, that if uniqueness is the only difference then it would be more natural to interpret 'The Φ is Ψ ' as 'Exactly one Φ is Ψ '. So Russell could be accused of leaving out some details, and it's in the details that trouble likes to hide. But this way of objecting is not my main concern either.

Russell shows how his theory can be used to explain three 'puzzling' observations about the way definite descriptions are used. But it has been claimed that there are other observations about the way they are used that not only can he not explain, but that go against what his theory predicts. These are the objections that concern me. I will consider two.

Strawson (1950, p. 330) pointed out that sometimes we are reluctant to give a truth value to a statement that is plainly false according to Russell's theory. Suppose someone were to announce 'The king of France is wise'. According to Russell he is saying that 'There is at least one king of France, there is at most one king of France, and whatever is king of France is wise'. Because the first conjunct is false the whole statement is false. But when asked whether the statement is true or false many people say 'neither, because there is no king of France'.

This is a problem for Russell, made worse by Strawson's intuitively appealing and incompatible account. Against Russell, Strawson says that the task of a definite description is, indeed, to refer to the object that it denotes. If there is no such object then it fails to refer. Furthermore, the task of a statement is to be true or false. If it contains a definite description that fails to refer, then the statement, in turn, fails to be true or false. So a statement can only be true or false if all of the definite descriptions that it contains denote an object. Nevertheless, a statement can still be *meaningful* even though it contains a definite description that denotes nothing. For it to be meaningful, says Strawson, there need only be *possible* (not necessarily actual) circumstances in which the definite description denotes an object and therefore in which the statement is either true or false. Finally, anyone who seriously makes a statement is taken to believe that what he says is true. So if the statement contains a definite description, then he is taken to believe that it denotes an object, because if he didn't believe that then he couldn't believe that the statement is true. Thus, definite descriptions are used on the *presupposition* that they denote. Strawson can explain our reaction to 'The king of France is wise' quite simply: We say 'neither' to the question of whether it's true or false because the statement fails to be either. We add 'because there is no king of France' to indicate to the speaker that his presupposition is false.

Strawson and Russell cannot both be right - one says that the task of a definite description is to refer and the other says that it's not. The current example seems to give Strawson the upper hand - our behaviour is explained by Strawson's theory, but not by Russell's. Worse still, it is *contrary* to Russell's. I don't want to take sides. What I *do* want to do is argue that this example does not threaten Russell as obviously as it might seem.

I think it's possible that our reaction to 'The king of France is wise' is indeed compatible with Russell's theory. Recall that according to him, this says that 'There is at least one king of France, there is at most one king of France, and whatever is king of France is wise', or, to keep things a bit simpler, that 'There is exactly one king of France, and he is wise'. This is a conjunction of two statements. If I think that either of the conjuncts is false then I will think that the conjunction as a whole is false. But if I just say that the conjunction is false I would not be conveying as much information as I could - I would not be conveying which conjunct it is that I think is false. To do that, I would have to add an extra bit. I might say, "'There is exactly one king of France, and he is wise' is false, because there is not exactly one king of France". Or I might say, "'There is exactly one king of France, and he is wise' is false, because he is not wise". If whenever such a conjunction is false it is *usually* because the second conjunct is false, then it would be handy to have a convention to the effect that when I declare the conjunction false it is taken to be because the second conjunct is false, unless I specify otherwise. Well, maybe for definite descriptions there is such a convention - that "'The Φ is Ψ ' is false" means by convention 'There is exactly one Φ and it is not Ψ '.

If so, it would explain how the puzzle about the law of the excluded middle arose - because of the convention the negation of 'The Φ is Ψ ' is taken to be 'The Φ is not Ψ ', rather than the logically weaker 'It is not the case that (the Φ is Ψ)' as it should be. It would also explain our reaction to 'The king of France is wise'. This cannot be true, because there is no king of France. But to say that it's false would be to say, according to the convention, that 'The king of France is not wise'. But that can't be true either, again because there is no king of France. So we can't say it's true, and we can't say it's false, so we say 'neither'. Unless we say 'False', and add 'because there's no king of France' to indicate that we have broken the convention. And this is what some people do in fact say. Is it strange to suggest that there might be such a convention? I don't think so, because it wouldn't be the only one like it. Consider the word 'but'. Logically, it behaves like the word 'and': ' Φ but Ψ ' is true if and only if ' Φ and Ψ ' is true. But, by convention, ' Φ but Ψ ' says a bit more than ' Φ and Ψ ' - it also says that Φ and Ψ are not normally true together. Because of this convention, we might think that ' Φ but Ψ ' is true, and yet be reluctant to say so (it would be strange to say, for example, that 'It's sunny but it's warm'). I think a plausible case can be made for the claim that definite descriptions are governed by a convention in the way described, and therefore that Strawson's objection to Russell is not as strong as it might appear.

The second objection I want to consider was raised by Donnellan (1966, 1970). He pointed out that sometimes a statement that Russell's theory says is false is actually judged to be true. He gives an example similar to this: We are at a party. You say to me 'the man drinking a martini is rather tall'. I know that the man you are talking about is actually drinking water from a martini glass, but I respond anyway with 'true'. According to Russell you have asserted that 'at least one man is drinking a martini, at most one man is drinking a martini, and whoever is drinking a martini is rather tall', which is false because the first conjunct is false. Moreover, I *know* that it's false, and so that's what my judgment should be. But it's not - it's 'true'. Trouble for Russell.

Donnellan claims that Russell is wrong to say that definite descriptions are *never* used to refer, because here is a case in which one is. He says that your statement succeeds - you have made the desired claim about the desired person - and so your use of the

definite description must have succeeded as well. But its success cannot have been due to correctly claiming the existence of something with certain properties, because there is nothing with the properties that it attributes. It can only be because it *picked out* the right person. So it must have been used to refer. Again, Donnellan and Russell cannot both be right, because one says that definite descriptions are sometimes used to refer and the other says that they never are. Again, I don't want to take sides, but just argue that this example does not obviously count against Russell either.

Donnellan claims that definite descriptions are sometimes used referentially, as in the example above, and sometimes used 'attributively', in the way that Russell describes. Moreover, the very same definite description can be used in either way. Consider a slightly different example. The chairman of the local Teetotalers Union has just been woken from his sleep and informed that a man is drinking a martini at the party. He calls me and asks 'who is the man drinking a martini'? Here, Donnellan says, the definite description is being used attributively, because it is all-important to the question having an answer that there actually is a man drinking a martini. Because there is no such person, I cannot answer either 'true' or 'false' to the chairman's question.

There is an important difference between the circumstances in which the definite description is being used in these two examples. In the first, you are in a position to *point* or *look* at the man you intend the definite description to denote. In the second, the chairman is not. I think it's open for Russell to respond as follows. In the first case, because you could look at or point to the man you didn't need to include his position in your definite description. You didn't need to include 'over there' or 'standing next to the heater' or anything like that. But they were included *non-verbally* in your description, and when I think about what you mean I add it to your description. I notice that nothing perfectly fits the description 'The man over there drinking a martini', but that something does perfectly fit the description 'The man over there who appears to be drinking a martini'. I judge that you'd be happy to use this description instead, and charitably put it in place of the one you actually used. So when I say 'true', I'm actually judging the statement 'The man over there who appears to be drinking a martini is rather tall'. And this statement is indeed true according to Russell's theory. In the second case, when the chairman calls me on the phone, I cannot charitably revise his definite description because nothing non-verbal is included in it. All I can go on is what he *says* - the words that he actually uses. That's why I cannot conclusively answer his question.⁴

I think that every one of Donnellan's examples of the 'referential' use of a definite description can be seen in the same way - as a case of someone using the principle of charity and non-verbal cues to modify the definite description to one that actually denotes something. Thus, when it appears that the judgment 'true' is being made of a statement that is false according to Russell, it is actually being made of a slightly different statement that is true according to Russell. So I think that the examples Donnellan raises are not so obviously a threat to Russell's theory as they might at first appear.

⁴ If your act of saying 'the man drinking a martini is rather tall' does implicitly include the phrase 'that' or 'over there', then Russell might actually agree with Donnellan that you are referring - not because you have used a definite description, but because you have used what he calls a 'logically proper name' (e.g. 'that') whose task is indeed to refer. So not only might this response argue away the apparent counter-example, but it might show that Donnellan's so-called 'referential' use of a definite description is really a combination of a Russellian attributive use with a Russellian logically proper name.

That concludes my account of Russell's theory of definite descriptions. I hope to have shown that he had some good reasons for developing it, that it can explain and predict some common ways that we use definite descriptions, and that some of its apparent faults can be argued not to be so.

REFERENCES

DONNELLAN, K. (1966), "Reference and Definite Descriptions", *Philosophical Review*, **77**, pp. 281-304.

—— (1970), "Proper Names and Identifying Descriptions", *Synthese*, **21**, pp. 335-58.

FREGE, G. (1892), "On Sinn and Bedeutung", *Zeitschrift für Philosophie und philosophische Kritik*, **100**, pp. 25-50; in M. Beaney (ed.), *The Frege Reader* (Blackwell, 1997), pp. 151-171.

MEINONG, A. (1904), "The theory of Objects", in *Untersuchungen zur Gegenstandstheorie und Psychologie*, Leipzig.

RUSSELL, B. (1905), "On Denoting", *Mind*, **14**, pp. 479-93; in R. C. Marsh (ed.), *Logic and Knowledge* (London, 1956), pp. 41-56.

—— (1919), *Introduction to Mathematical Philosophy*, London, chapter 16.

STRAWSON, P. (1950), "On Referring", *Mind*, **59**, pp. 320-44.