WHAT REFERS? HOW?

Alex Barber

I.   Introduction: what refers?

Guns don't kill people, people kill people. This National Rifle Association slogan has a popular analogue in the philosophy of language: *Words don't refer to things, people refer to things*. Thus we have Peter Strawson asserting, in his reply to Bertrand Russell, that:

> 'referring'… is not something an expression does; it is something that someone can use an expression to do. (Strawson 1950: 326)

Echoing him, Leonard Linsky writes:

> Of the first importance [in discussions of the reference relation] is the consideration that it is the users of language who refer … and not, except in a derivative sense, the expression which they use in so doing. (Linsky 1963: 74)

Noam Chomsky cites Strawson in support of a similar-sounding view:

The question, "to what does the word X refer?," has no clear sense… . In general, a word, even of the simplest kind, does not pick out an entity of the world, or of our "belief space"… . (Chomsky 2000: 181)

James McGilvray, too, criticizes those who:

…think of reference as a conventional relationship between word and world, or perhaps even look for a "natural" relationship in a realist construal of information theory; in either case, they apparently ignore the fact that people use words (and they use them in many ways) to refer to things. (McGilvray 2005: 15; see also McGilvray 1998)

But just as campaigners for gun control persist in talking of the lethal properties of guns, many philosophers of language, including W. V. Quine and Nathan Salmon, continue to talk of the referential properties of words, conceding only that the identity of these properties must sometimes be relativized to a context of utterance.

One thinks of reference, first and foremost, as relating names and other singular terms to their objects. (Quine 1990: 27)

[T]he semantic attributes of expressions are not conceptually derivative of the sppech acts performed by their utterers, and are [instead] intrinsic to the expressions themselves, or to the expressions *as* expressions of a particular language and as occurring in a particular context. (Salmon 2004: 238)

And of course, formal semantics, even when cognitivist in orientation, involves the ascription of referential properties to expressions:

> 26 a.   *Rabbit* refers to rabbits.
>
> ….
>
> Proper nouns like *Jill* … are analysed as having persons as their semantic values, that is, as referring to those individuals:
>
> 9 a.   Val(x, *Jill*) iff x = Jill. (Larson & Segal 1995: 41, 118)

So what refers, words or people? Perhaps both do. Many, including Saul Kripke and Kent Bach, have insisted on the dissociability of two distinct notions of reference:

> The speaker's referent is the thing the speaker referred to by the designator, though it may not be the referent of the designator, in his idiolect. (Kripke 1977: 238)

> Failing to distinguish speaker reference from linguistic reference … inevitably leads to theoretical confusion. (Bach 1987: 6)

Completing the logical space, a fourth possibility is that neither words nor people refer.

Rather than launching into an investigation of which of these four possible views is correct, it would be as well to ask whether and why it matters how we fill the subject position when using the verb 'to refer'. I take this question up in section II. The rest of the paper uncovers then tackles a specific puzzle about the relation between speaker reference and word reference. The ingredients of the puzzle are entirely familiar, having to do with the apparent intentionality (in senses loosely associated with Brentano) of speaker reference, a trait that

word reference seems to lack. In view of this apparent contrast, how could both speakers and words refer? The puzzle is introduced in III and firmed up in sections IV-VIII, with a solution offered in IX. Part of the firming up process involves explaining how some standard views on speaker reference and word reference seem to have led to the puzzle being either overlooked or dismissed.

II.   What turns on what refers?

The question of whether it is speakers or words or both or neither that refer could be – and clearly often is, tacitly at least – approached naïvely, as a question about proper usage of the verb 'to refer'. But why should we care about the folk's view on what refers? Folk views sometimes constitute useful if crude proto-theories, platforms from which to develop more rigorous proposals. That does not seem to be the case here. But in any case, the folk's opinion is unclear. Asking explicitly yields no helpful answer, since the folk are vague in their views and easily led away from their initial answer no matter what it may have been. Notice in addition that acceptability judgements vary from one linguistic culture to another. As readily as English speakers allow that universities educate students, French speakers allow that lawnmowers mow lawns, even though this latter is unacceptable to most English ears. Unless we aspire only to a parochial theory of reference, there is little to be gained from asking English speakers which, if any, "sounds better" out of the views of Quine and Strawson, quoted above.[1]

Rather than addressing a semi-interesting question about ordinary usage with no clear answer, I will attempt to raise a genuine puzzle about reference by coming at the topic of opacity from an unfamiliar angle. Others have had reasons of their own for taking a view on whether it is words or speakers that refer. I will summarize two such reasons briefly, below, not because anything I say bears directly on them or because there are no other reasons a

person could have,[2] but because they impinge, obliquely, on what I have to say about the puzzle that is my more immediate concern.

The essence of Strawson's point is that if we treat referring as something people can do with words, Russell's theory of descriptions and its corollary, his theory of ordinary names, have no purpose. Whenever a person attempts to refer using definite description (perhaps disguised as a name), they make, or affect to make, certain presuppositions – that there is a unique contextually salient satisfier of the description, for example. They succeed in referring only if these presuppositions are met. Russell's mistake is to confuse the presupposition relation, which holds between a *speaker* and some condition, for an entailment relation, i.e. one that holds between the *expression used* and this same condition. His theory of the logical form of descriptions explains this entailment, but if Strawson is right, no such explanation is needed. One can utter a definite description in an attempt to refer but fail to refer because nothing uniquely satisfies the description. This is no more mysterious than the fact that one can, for example, swerve in an attempt to avoid an obstacle but fail because there was no obstacle to avoid. Neither calls for a semantic explanation.[3]

If referring is indeed something people do with words and not a standing characteristic of words, a more general moral one might draw is that attempts to deliver a systematic theory of the referential properties of the expressions of a language are misconceived. Referring would be an aspect of the use of language rather than of a stable body of knowledge codified in the language faculty – an aspect of performance rather than competence as Chomsky would once have put it. Ambitions to provide a theory of this use would be as hopeless as behaviourist attempts to capture our dispositions to produce particular sounds. Intuitions about reference may feed into the explanation of systematic, relatively contained phenomena, such as binding relations. But according to Chomsky, McGilvray, and others (e.g. Pietroski 2003), attempting to construct a theory of reference as philosophers of language tend to use

this phrase, in which word-world relations entail a compositional theory of truth for the sentences of a language, is the wrong way to set about developing a natural science of language.

I will remain as neutral as I can on both Strawson's and Chomsky's concerns, which are broader than my own. But both are connected, tangentially, to the puzzle to which I now turn.

III. A puzzle about reference

I am interested in the ramifications of an apparent contrast between the characteristics of word reference, assuming there is such a thing, and the characteristics of speaker reference, again assuming there is such a thing. Given assumptions I will come to, speaker reference is arguably intentional in the Brentano sense. That is to say, facts about speaker reference seem to respect neither of two logical laws, the law of existential generalization and the law identity. Were it a fact that someone had touched Santa Claus, it would follow that Santa Claus existed; and from the fact that Hesperus is a planet it would follow that Phosphorus is a planet. But from the fact that a person has referred to Santa Claus it does not follow that Santa Claus exists; and from the fact that a person has referred to Hesperus it does not follow that they have referred to Phosphorus. Word reference, by contrast, and once again given assumptions I will come to, is arguably intentional in neither respect: a word cannot refer to Santa Claus given that no such person exists, and if a word refers to Hesperus it refers equally to Phosphorus.

To keep things manageable I will focus exclusively on the failure of the law of identity until section IX, at which point I will show that a solution to the version of the puzzle thrown up by substitution failure works to solve the version of the puzzle thrown up by failure of existential generalization. In the meantime, and in the form of it that I will be focussing on, the puzzle arises out of a tension between two claims, I and II below.

I. Speaker reference is substitution unfriendly.

II. Word reference is substitution friendly.

A property (or perhaps, a "property") is substitution friendly if and only if it always respects the law of identity, so that if some object has that property then anything identical with the object also has the property. I and II, then, make contrasting claims about whether being referred to is among those properties that respect the law of identity: if it is a speaker doing the referring, the law is not respected; if it is an expression doing the referring, the law is respected. If Jane refers to Hesperus she has, according to I, not necessarily referred to Phosphorus, despite their being one and the same planet. By contrast, if 'Venus' refers to Hesperus then according to II it must also refer to Phosphorus.

Let me stress that the puzzle is *not* simply that I and II are immediately plausible but inconsistent with one another. Rather, it is that further claims, which are plausible, collectively entail I and II; and that another claim, also plausible, is provably inconsistent with I and II. Programmatically, the dialectic is as in Figure 1.
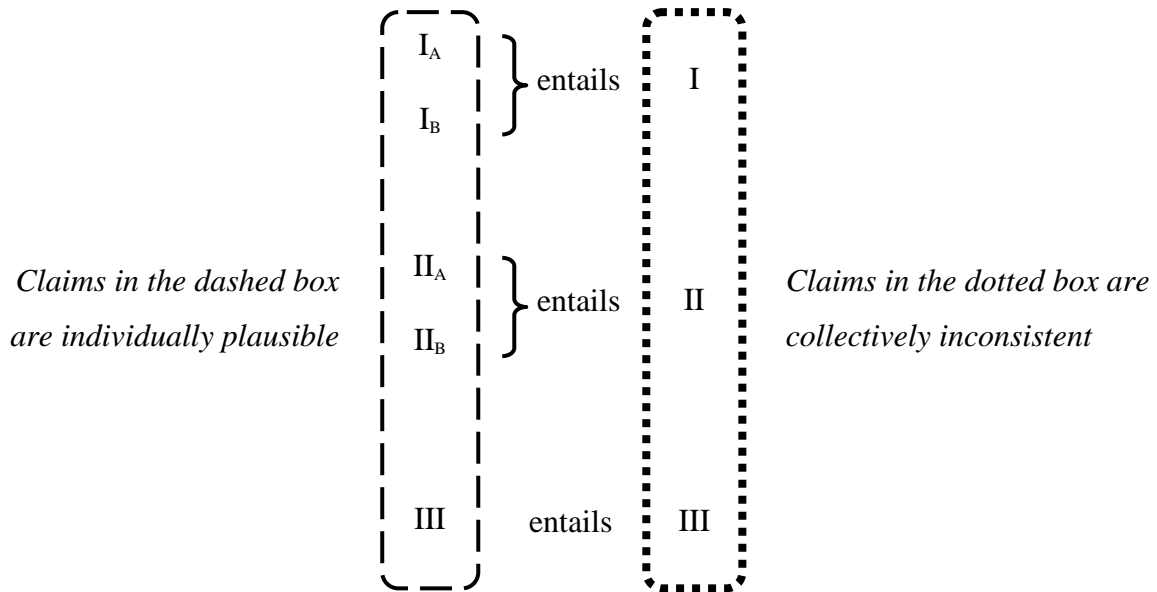
Figure 1. The structure of the puzzle.

The puzzle comes to this: five plausible claims – I_A, I_B, II_A, II_B, and III, none yet stated – cannot all be true. I and II merely play a central role in demonstrating that the five claims are inconsistent with one another.

In this section I will be concerned only with the entailment of I by I_A and I_B and the entailment of II by II_A and II_B (and of course with saying what I_A through to II_B are). Only in the sections that follow will I be concerned with the alleged inconsistency – arguing for its seriousness, and suggesting a solution.

Assumption I_A is that what a speaker refers to in uttering a word is constitutively determined by that speaker's intention in performing the utterance. There are various forms the relevant intention could take. In what follows, I will take I_A to be the schematic generalization[4] of which the following is an instance:

Jane uses 'Venus' to refer to Hesperus if and only if her intention in uttering the word is that her audience should think of her doing so as a reference to Hesperus.

8

The precise formulation of the intention is not critical. I have adopted this one rather than, say, '…her intention in uttering the word is to refer to Hesperus' because intentions, in general, must be liable to being acted on unsuccessfully.[5] Jane could easily fail in her intention to be thought of as referring to Hesperus, but if the intention that constituted her referring to Hesperus was the intention to refer to Hesperus, her intention would be self-fulfilling. For all that, there is an unusual connection between the success condition for her intention and her intention's content: to succeed in bringing about what she intends, it is both necessary and sufficient that her intention be recognized. In this respect, referring can be grouped along with promising, asking, insinuating, and other meaningful acts like depicting, and distinguished from other kinds of non-meaningful acts such as kicking or finding. Arguably it is a mark of meaningful acts that they belong to some kind k such that one counts as having $k$ed if and only if one has acted with the intention of being thought of as having $k$ed (where additional characteristics are needed to distinguish meaningful act types from one another). I_A merely combines this thought about what it is to be meaningful with the platitude that for a person to refer is for them to perform a meaningful act.

I_B is a claim about all intentions, not only semantic ones. Suppose Jane intended to pat Rex. It does not follow that she intended to pat Fido, even if Rex and Fido are one and the same dog. It may be that she had no idea that Rex and Fido were the same dog. I_B, the generalization of this, is the claim that intending is substitution unfriendly. Of course, not everyone agrees that psychological attitudes are substitution unfriendly. I discuss their views in section VIII. At this stage I am merely attempting to make good the dialectic of Figure 1, and it is enough that many would find I_B plausible. If we apply I_B to the intentions found in I_A, I follows: speaker reference is substitution-unfriendly. This result is not peculiar to my idiosyncratic formulation of the intention that constitutes someone having referred to

9

something. The same conclusion is invited by any formulation instances of which name or idiosyncratically describe the referent inside the scope of 'intends' or some other psychological operator on the right-hand side of the 'iff'. This or something approaching it appears to be a commitment of many existing definitions of speaker reference or related notions (see e.g. Kripke 1977: 238; Evans 1982: 318-20; Bach 1987:52; Barker 2004: 113; Davis 2005: 162, 202-3; see also Searle 1969: 94 and the penultimate note of this paper).

What about II? Before searching for plausible claims that entail it, it is important to distinguish it from a similar sounding but more familiar thesis. Let II' be the claim that licences any inference with the same pattern as the following:

'Hesperus' refers to Venus.

'Phosphorus' refers to Venus.

∴ '…Hesperus…' is true if and only if '…Phosphorus…'.

In other words, according to II', co-referring expressions of the object language can always be substituted *salva veritate* within sentences of the object language. (There are, of course, some widely debated apparent exceptions to this pattern, including attitude reports.) By contrast, the inference licensed by II is:

'Venus' refers to Hesperus.

Hesperus is Phosphorus.

∴ 'Venus' refers to Phosphorus.

II' insists that substitution of one object-language expression for another co-referring object-language expression does not affect the truth of the object-language sentence within which the

substitution is made. II concerns the choice between two co-referring expressions of the metalanguage. The two claims are related since II asserts, in effect, that constructing sentences using '"Venus" refers to…' does not generate exceptions to II'; but II and II' are distinct for all that.

That word reference is substitution friendly follows from II_A and II_B. II_A is the claim that the truth-conditional properties of sentences lacking opaque operators are substitution friendly. Given that Hesperus is Phosphorus, for example, if 'Venus is wet' is true if and only if Hesperus is wet, then it is true if and only if Phosphorus is wet. II_B is the converse of compositionality. Compositionality (in the sense I am using the expression here) is the thesis that there is a function from the referential properties of simple expressions onto the referential properties of the complex expressions built out of them. As a special case of this, if, for each word within a sentence, the condition is set for an arbitrary entity being that word's referent, then there is no room for manoeuvre on what it would take for that sentence to be true. The converse of this, II_B, is that once the condition for a particular sentence being true is set, there is no room for manoeuvre on the condition for any expression within it to refer to something: change the requirement an object must meet for it to be the referent of any word in a given sentence and you will change the range of circumstances in which that sentence is true. For example, if a name 'Nelson' referred not to Nelson but to Churchill, the circumstances under which 'Nelson was British' is true would change.

The proof of II from II_A and II_B, summarized below, is essentially a matter of showing that the substitution friendliness of reference conditions follows from the substitution friendliness of truth conditions, given the converse of compositionality.

II_A If Hesperus is Phosphorus then the biconditional, that 'Venus is wet' is true-iff-Hesperus is wet iff 'Venus is wet' is true-iff-Phosphorus is wet, holds.

II$_B$ The biconditional, that 'Venus is wet' is true-iff-Hesperus is wet iff it is also true-iff-Phosphorus is wet, requires the biconditional, that 'Venus' refers to Hesperus iff it refers to Phosphorus.

∴   If Hesperus is Phosphorus then 'Venus' refers to Hesperus iff it refers to Phosphorus.

The conclusion, here, is but an instance of II, but the proof clearly generalizes.[6]

Against this supposed demonstration of II one might ask whether the condition for the truth of 'Venus is wet' really can be expressed using 'Phosphorus' in place of 'Hesperus' (assuming use of the latter were acceptable). After all, the resulting expression of the condition differs in meaning, or cognitive significance, from the original one. I will reject this response in detail in a later section. At present I am concerned only with making a prima facie case for the existence of a puzzle. All the same I will make two related points even here. First, if a good case can be made for thinking that meaning, or cognitive significance, is substitution unfriendly, then that does not necessarily represent a problem for the view that truth conditions are substitution friendly, which is all that II$_A$ demands. It might equally be interpreted as a challenge for the simple identification of meaning with truth conditions. While meaning and truth conditions are intimately related, their relation is not well captured by the slogan that the meaning of a sentence *is* its truth condition. Second, there is good reason to think that the condition for a sentence's being true is substitution friendly. Were it not, we would be forced to re-evaluate our endorsement of the law of identity and the hypothetical syllogism in contexts that seem entirely benign, such as the two inferences below:

12

Hesperus is Phosphorus

∴   Hesperus is wet iff Phosphorus is wet

p iff Hesperus is wet

Hesperus is wet iff Phosphorus is wet

∴   *p* iff Phosphorus is wet

But if these inferences are legitimate – assuming any grammatical replacement of '*p*' in the second inference – they can be concatenated to legitimize our illustrative instance of II$_A$ (as the more detailed proof in the previous note makes clear):

'Venus is wet' is true iff Hesperus is wet

Hesperus = Phosphorus

∴   'Venus is wet' is true iff Phosphorus is wet

Unhappiness with this concatenation is likely to spring from claims about what attributions of truth conditions are meant to achieve. It may be that they achieve this using 'Hesperus' but not 'Phosphorus'. This unhappiness, which springs from a desire to maintain a tight connection between truth conditions and meaning, is taken up in section VII.

IV.   Interlude: summary so far and sketch of what follows

I have shown that two claims, I and II, which are in tension though not inconsistent with one another, are entailed by four other claims, I$_A$, I$_B$, II$_A$, and II$_B$, all of which have some claim to plausibility. I will now consider some different ways in which one might be underwhelmed by

this tension. Each way corresponds to some stance one might take to the question of whether it is words or speakers that refer. All the stances I look at give rise either to resistance to my suggestion that there is something to worry about in the alleged tension between I and II, or to a failure even to acknowledge this tension.

First, a person may agree that there are two notions but insist that, for this very reason, there is no tension. I call this position *dualism*. By not conflating the two kinds of reference, e.g. by abandoning the ambiguous expression 'reference' entirely and replacing it with two terms, one for word reference and another for speaker reference, the so-called tension evaporates. Or so dualists will claim. In the next section I argue against dualism by showing how a third claim, the 'III' of Figure 1, generates an inconsistency between I and II even if one keeps the two notions of reference distinct.

In contrast, *monists* maintain that there is in fact only one notion of reference, the existence of distinct notions being a readily explicable illusion. If there is only one notion of reference, there cannot be a tension between the "two" notions. I respond to monism in section VI, first by arguing for its implausibility, and second by showing how the inconsistency arises even for monists, albeit in a different form.

A third stance, *cognitivism*, is the topic of section VII. Cognitivists assume that attributing referential properties to words is best construed as describing an aspect of a speaker's psychology. If word reference is psychological in this way, it could be as substitution unfriendly as speaker reference, and for the same reason.

The final source of hostility to seeing anything worth labelling a "puzzle" in the tension between I and II takes the form of a disjunctive dilemma. Names refer either directly or indirectly. If they refer directly, the tension can be explained by rejecting I_B, the claim that psychological attitudes are substitution unfriendly. If they refer indirectly, via a description that is associated somehow with the name, then other aspects of the argument for the

14

inconsistency become vulnerable. So while we may not know which of the two theories of names is correct, the tension is a problem for no one. I consider this disjunctive response in section VIII.

If I am right that these attempts to trivialize the puzzle all fail, the five propositions are inconsistent and we must abandon at least one of them. In section IX I explore the possibility that the source of the problem lies with intention-based accounts of speaker reference. I do not deny that an acceptable intention-based account of reference is available, but its form must be such as to allow that speaker reference is substitution friendly even if intentions are not. I go on to suggest such a form.

V.   Dualism about reference: two notions, not one?

I and II taken together sound vaguely paradoxical, but could equally serve as a further reminder of the importance Kripke, Bach, and others attach to the distinction between speaker reference and word reference. An effective way of maintaining the distinction would be to introduce different terms, such as 'picking out' for what speakers do and 'designating' for what words do. This distinction is a real one in at least this: speaker reference (or picking out) and word reference (or designating) are dissociable. I may refer to an overbearing sub-Dean as 'Napoleon', using a word that designates a dead French military leader to pick out a living British civilian. Or I may accidentally refer to someone using the name of someone else who I had been thinking about moments earlier.

The two notions may be distinct, but a puzzle arises anyway if the two notions coincide in what I will call normal circumstances. (I will say much more about "normality" shortly.) That is, I and II are inconsistent if taken together with III, which is the claim that there is normally a symmetry between speaker reference and word reference.

III.   *"The symmetry principle":* In normal circumstances, a speaker refers to whatever

the word she or he uses refers to.

In other words, for a speaker to refer to something, it is normally both enough and required that the word used refers to that thing. If I use the name 'The Admiral Lord Nelson', for example, then barring performance error, irony, massive ignorance of early 19th century European history, and other arguably abnormal factors, I will have referred, not to Napoleon, but to Nelson. It is both enough and necessary for me to refer to Nelson that I use a word that refers to Nelson. This symmetry is inconsistent with the conjunction of I and II:

The inconsistency of I, II, and III

Suppose Jane uses 'Venus' to refer to Hesperus in what, let us suppose, are normal circumstances. Did she refer to Phosphorus? The symmetry principle (in one direction) tells us that the word she used refers to Hesperus, and hence, by II, that it refers to Phosphorus, and hence, by the symmetry principle (in the other direction), that Jane did indeed refer to Phosphorus. This shows us that, in normal circumstances anyway, speaker reference is substitution friendly, *pace* I.

Whatever the merits of this argument against dualism, it survives the replacement of 'Jane refers' and '"Venus" refers' by 'Jane picks out' and '"Venus" designates'. Reservations are likely to turn, rather, on the qualification in III: 'In normal circumstances,…'.

The most immediate concern is that the inconsistency shown above is not between II, III, and I, but between II, III, and the claim that substitution failure in speaker reference can occur in normal circumstances. This leaves open the possibility that speaker reference is substitution unfriendly, but only in abnormal circumstances. Were that so, the tension

Hesperus-Phosphorus cases give rise to could be assimilated to the class of cases in which it is known that speaker reference and word reference fail to ride in tandem.

Two ways of responding to scepticism about the 'in normal circumstances' qualification in the anti-dualist argument are available. The first, and probably more promising, would be to formulate the demonstration of an inconsistency without using the qualification, perhaps by replacing III with something else. I do precisely this in the next section, in the course of looking at monism. The second would be to impose more robust constraints on the qualification instead of leaving it so vague. These constraints would need to be tight enough to render the use of 'Napoleon' to refer to a sub-Dean abnormal but liberal enough for Hesperus-Phosphorus cases to be classified as normal.

Swift's *Gulliver's Travels* contains a clue to how this second option might be pursued. Gulliver's host during his visit to the Grand Academy of Lagado parades before him what are in fact parodies of various 18th century scientific and philosophical innovations. Among them is a progressive new language, innovative for its lack of words. Gulliver is informed in its defence that:

> [s]ince Words are only names for Things, it would be more convenient for all Men to carry about them, such Things as were necessary to express the particular business they are to discourse on. (Swift 1726: 76)

'Speakers' of this language are weighed down with the objects they wish to 'talk' about with their colleagues in silent 'conversations'. The moral of Swift's satire is that language's value depends on its enabling us to share thoughts about a thing without being contextually tied to that thing. By using a word in a thing's stead we need not carry the thing itself around – and need not rely on other heavily contextual cues to enable our interlocutor to figure out what we

are referring to. Of course, we *can* still communicate in a way that does rely on heavily contextual cues (including presentation of the referent as a limiting case). When this happens, or when the audience must diagnose a performance error in order to identify the referent, a speaker may use a word to refer to something other than what the word refers to. But the use of a word to refer is normal to the extent that exploiting heavily contextual cues and diagnosing performance error are not necessary for the audience to identify the referent. In these circumstances, the choice of word does the trick, which is in effect what III is claiming. Jane's using 'Venus' to refer to Hesperus in the course of asserting that Hesperus is wet appears to be normal in this (loose but tighter) sense.

A lingering worry, here, is that no reason has been given to think that III works in both directions. Suppose Jane uses 'Venus' to refer to Hesperus, in a situation normal enough that 'Venus' refers to Hesperus. By II, 'Venus' then also refers to Phosphorus. At this point someone could dig their heels in. Perhaps the situation counts as fully normal only if Jane knows that Hesperus is Phosphorus, and because she does not, she has not referred to Phosphorus. In less than fully normal circumstances the most we could say is that the word Jane uses refers to Phosphorus only if there is something identical to Phosphorus that she has referred to.

This possibility is hard to reconcile with any plausible account of word reference, without which dualism would lack plausibility in any case. I will say more about what the options are in the next section, but on a conventionalist account, for example, a word refers to that which people regularly use it to refer to. It must be possible for it to be used on different occasions and by different people to refer to the same thing known in different ways without sabotaging the potential for a convention. If S_1 uses W to refer to O_1 and O_2 without realising that O_1 and O_2 are O_3, S_2 uses W to refer to O_2 and O_3 without realising that O_2 and O_3 are O_1, and S_3 uses W to refer to O_3 and O_1 without realising that O_3 and O_1 are O_2, that could count

18

as the community of $S_{1...}S_3$ using W in a common enough way for the convention to establish itself. It would be peculiar for the convention itself to discriminate in favour of one particular speaker's perspective on the object. All objects admit of a multiplicity of perspectives, but why should one particular perspective be definitive of *the* use defining the convention. More plausible (in so far as conventionalism is plausible) would be that all uses from which word reference is extrapolated should be thought of as some person or other's use of the word to refer to the object that is taken up the referent of the word.[7]

VI. Monism about reference: one notion, not two?

If there is only one kind of reference, be it word reference or speaker reference, there can hardly be a tension between the contrasting characteristics of two notions of reference, let alone a proof along the lines of the previous section of an outright inconsistency. The quotations from Quine and Strawson in section I represent opposite strands of this reaction to the puzzle. Quine's view is less widely supported today. Those who accept that a given word refers to something tend to require that its doing so depends in some way on people using it or having used it to refer to that thing, though perhaps not solely on this. Strawson's view is initially more plausible. According to him and Linsky (and see also Bach 1987: 40), words don't refer at all, save in an entirely derivative sense, and the very question of which two-place reference relation is the fundamental one – i.e. the two-place relation between words and objects or the two-place relation between speakers and objects – is misconceived. Talk of words referring or of speakers referring, in each of which reference is superficially a two-place relation, distorts the nature of the three-place relation that reference really is, i.e. a matter of a speaker$_I$ using a word$_{II}$ to refer to an object$_{III}$.[8] The NRA slogan arguably exploits a parallel confusion: it is not that guns$_I$ kill people$_{II}$, or that people$_I$ kill people$_{II}$. Rather, people$_I$

19

use guns$_{II}$ to kill people$_{III}$. The use of distinct shorthands, in each of which the three-place relation appears to be two-place, creates the illusion of a substantial debate.[9]

In what follows I will, first, offer a reason for being suspicious of this form of monism. But for those who find this reason unpersuasive I will go on to present a monist-friendly version of the puzzle: an inconsistency similar to that of the previous section but reached without the assumption that words refer (or designate, etc.).

To begin, then, why is the view that only speakers refer implausible? The answer springs out of an intention-based account of speaker reference along the lines of I$_A$ plus two plausible psychological assumptions. If what a person refers to is a matter of what they intend to be thought of as referring to, they can refer only if they can form the relevant intention. But forming intentions requires an expectation of success. Right now, for example, I am not in a position to leap to the moon, and because I know this I cannot even form the intention to leap to the moon. To form that intention, I would need to believe that leaping to the moon is possible. I cannot simply choose to believe this, since belief formation is itself subject to a psychological requirement: that one have some evidence in favour of the proposition to be believed. These two principles – expectation of success as a requirement on intention formation, and evidence as a requirement on belief formation – taken together with I$_A$ mean that I can only refer if I have some reason to think that I will be recognized as having referred. In some circumstances this condition will be met even in the absence of any stable relation between word and referent. (That performance errors do not always sabotage communication is proof of this – for instance, someone may refer to someone called Paul using 'Peter' without the audience being misled as to who was referred to, perhaps because they can see the speaker looking straight at a known Paul, or because, even when this Paul is absent from the conversational arena, the speaker attributes to the person she calls 'Peter' a number of properties that could only plausibly be ascribed to the Paul.) But in other circumstances, those

I hesitatingly labelled 'normal' in the previous section, more is required. What this more comes can be left open here, since any spelling out is likely to yield a notion of word reference that is distinct from speaker reference. A traditional understanding of the extra requirement is that of a convention relating the word's sound to the speaker's referent. The existence of such a convention would, in that case, constitute the word's having the referent it does. Others, sceptical of the explanatory potential of conventions, assume instead merely a stable association in the speaker's psychology between the word they use and the object they refer to (see Laurence 1996, Mercier 2003). What constitutes a word's referring to what it does refer to will, in that case, be defined in terms of such stable associations. Conventionalism and cognitivism (as I will call the second view) both require that, in normal contexts at least, speaker's can refer only if words refer, contrary to monism

A critic of this argument could point to live controversy surrounding both psychological assumptions. Against the expectation requirement on intention formation, for example, it does not seem implausible that many first-world war soldiers who fired their decrepit rifles across no-man's-land could have done so with the intention of hitting someone, despite their having little reason to expect success. Structurally similar cases involving lottery tickets have given rise to the so-called lottery paradox. There are linguistic instances, too: addressing people in English while travelling outside English-speaking countries can have a vanishingly small chance of success, but one often attempts it before resorting to arm-waving. On the other hand there does seem to be a psychological block to forming the intention to jump to the moon, and an obvious sense in which such a block would be adaptive. There is also a clear selective advantage to forming only beliefs for which one has evidence.[10] This has not prevented some – doxastic voluntarists – from rejecting the second psychological assumption used in the argument above. Here is no place to try to settle these disputes. I will therefore only reiterate that both are highly plausible for many cases and have clear selective advantages, and express

the hope that troublesome cases can be assimilated in a way that does not threaten their use in the argument against monism.[11]

A second line of response to the anti-monist argument springs from the thought that word reference is ultimately definable in terms of established usage, or beliefs about established usage or potential usability, or something along those lines. If some such definition is available, so that the notion of word reference can be reduced to that of speaker reference, can't we say that there is only one notion after all? The difficulty with this response is that, even if the inconsistency does not threaten the very prospects for a reduction, the fact that there is only one *fundamental* notion of reference – in the sense that all other relevant notions of reference reduce to it – does not mean there is only one notion of reference. Average height in a population can be defined in terms of, and hence reduces to, the notion of the height of an individual. It does not follow that they are the same notion. Likewise, that word reference could be defined in terms of speaker reference would not show that word reference has speaker reference's characteristics. It would still be possible in principle for that latter to be substitution friendly while the former is not.

A final reservation over the argument calls for a more serious response. Someone could insist that a motley conglomeration of mental states underpin our semantic abilities, including knowledge of the world, syntactic competence, and ad hoc on-the-hoof reactions to specific utterances. Nowhere in this *mélange* should we expect to find knowledge of, or belief in, a stable association between words and things. No contribution is made to the explanation of linguistic practice by such an association. On such a view, semantics as traditionally conceived – a substantial discipline sitting between syntax and pragmatics and dedicated to tracing word-world relations – withers away, along with the notion of word reference that is essential to the generation of the puzzle in the previous section. This is the view highlighted at the end of section II and associated with Chomsky among others (e.g. McGilvray 1998: 228).

For those sympathetic to this view, and indeed to anyone unhappy with the anti-monist argument just given, what follows is a monist-friendly version of the puzzle.

Even if words do not refer (in a sense that is even moderately invariant over contexts), allowing that speakers can use words to refer commits one to the mental equivalent of referring words. Following tradition, call these entities *mental representations*. The inconsistency reached in section V can be generated in a different form, couched in terms of the characteristics of the referential properties of mental representations rather than of words.

In expressing an opinion about something on a particular occasion I may use a word to refer to it. I may never have used, and never again use, this same word to this same referential end, and it may be that no one else uses this word to refer to this thing, either. Still, I will be able to do this only if I have a mental representation of the entity I have referred to. My utterance of the word *invokes* this mental representation, which is to say that the mental representation is tokened as part of the intention that constitutes the speaker's having referred using the word. This mental representation has referential properties of its own, a fact that is intimately connected to my ability to refer when I use a word that invokes it. This intimate connection can be captured in III*, a modified version of III.

III*. *"The symmetry principle*"*:* When a speaker refers using a word, she refers to whatever is referred to by the mental representation invoked by her utterance of that word.

That is, for a speaker to refer to something, it is enough and it is required that the word used invokes a mental representation referring to that thing. Using this to reach an inconsistency between speaker reference and mental-representation reference, requires updating II into II*:

23

II*.   Mental-representation reference is substitution-friendly.

In other words, if Jane's mental representation refers to Hesperus, then it also refers to Phosphorus given that Hesperus is Phosphorus – whether Jane realizes this or not.

II* follows from suitably modified versions of II_A and II_B. II_A held that the truth conditions of sentences are substitution friendly. II_A* holds that complex mental representations with truth conditions are substitution friendly. Given that Hesperus is Phosphorus, if p, a mental representation, is true iff Hesperus is wet, then it is also true iff Phosphorus is wet. Denial of this would be hard to reconcile with the congruence of satisfaction conditions across a wide range of attitudes, including believing, desiring, and intending. It may be possible to hold the belief that polenta is yellow without holding the belief that cornmeal is yellow, but given that polenta *is* cornmeal, the belief that polenta is yellow will be correct if and only if the belief that cornmeal is yellow is correct. Similarly, while it may be possible to intend to pat Fido without intending to pat Rex, the fact that Rex is Fido means that any intention to pat Fido will be successfully acted upon if, and only if, the intention to pat Rex is successfully acted upon. Finally, while it may be possible to desire that Superman takes you out on a date without desiring that Clark Kent takes you out on a date, the fact that Clark Kent is Superman means that the desire to be taken out on a date by Superman will be met when, and only when, the desire to be taken out on a date by Clark Kent is met.

II_B held that a difference in the reference conditions for a word gives rise to a difference in the truth conditions of any sentence in which the word appears. II_B* holds that a difference in the referential properties of a mental representation gives rise to a difference in the truth conditions of complex mental representations containing the first mental representation. This is highly plausible. For example, if mental representation C referred to Napoleon not Nelson,

a belief containing C would be correct according to whether something was true of Napoleon, not according to whether it was true of Nelson. The proof of II* from II$_A$* and II$_B$* is essentially the same as the one set out in section III.

Now, finally, we are in a position to prove the inconsistency:

The inconsistency of I, II*, and III*

Suppose that Jane uses 'Venus' to refer to Hesperus. Using II* and III*, it is possible to show, contrary to I (given the arbitrariness of the example), that she referred to Phosphorus. For in referring using 'Venus' she will have invoked a mental representation, call it *V*, that (by III*) refers to Hesperus. But from this and II* (and the fact that Hesperus is Phosphorus) it follows that *V* refers to Phosphorus. By III* (again) she has, therefore, referred to Phosphorus.

Jane may well have two mental representations, both of which refer to Hesperus/Phosphorus but only one of which is invoked by her utterance of 'Venus'. But this is irrelevant to the demonstration of an inconsistency. It challenges neither the premises nor any step of the argument based on them.

This new version of the puzzle improves on the old one in several respects. For a start, it makes no assumption that words refer. This was the initial motive for developing it. In addition, because it uses III* in place of III, the symmetry it presupposes is not confined, perhaps dubiously, to "normal circumstances". But then, the new version also introduces what some would see as weaknesses. In particular it assumes some version of the representational theory of mind. How robust or extensive its commitments are in this respect is hard to say, but it is only right to acknowledge in passing that a persistent and significant minority reject the

25

representationalist vision.[12] Which version of the puzzle one finds more compelling will therefore vary according to one's broader theoretical commitments.

VII. Cognitivism: word reference as a feature of speaker psychology?

I have looked at two takes on the relation between speaker reference and word reference – monism and dualism – and at how they disguise the very existence of the puzzle I am interested in. The third stance I wish to consider is really a view on word reference alone, but it, too, has the effect of hiding the puzzle. Contrary to II it appears to offer a route to treating word reference as substitution unfriendly, just as, according to I, speaker reference is.

According to cognitivists about word reference, claims about what a word refers to is properly to be thought of as an explicit statement of something known or believed by an individual speaker or group of speakers – or perhaps tacitly known or tacitly believed, since the knowledge may differ in one way or another from regular knowledge or belief. Abstracting from internecine controversies that are here unimportant, let us fix on 'tacit belief' as the label for the relevant cognitive relation, whatever its nature. On the cognitivist picture, overlap in what a group of language users tacitly believe about the referential properties of words, such as it is, contributes to the capacity each has to understand the others, such as it is. Talk of words referring *simpliciter* is loose, say cognitivists. It needs to be embedded in a psychological context. What words refer to independently of their being tacitly believed to refer, in so far as that makes any sense, is explanatorily otiose.

If word reference were substitution friendly in the same way that intentions are, the discrepancy between I and II, on which the puzzle trades, would evaporate. On the present construal of the ascription to words of referential properties, this promises to be the case. Suppose that Jane could tacitly believe the content of (1) without tacitly believing the content of (2). That is, suppose that (3) could hold without (4) holding.

(1) 'Venus' refers to Hesperus.

(2) 'Venus' refers to Phosphorus.

(3) Jane tacitly believes that 'Venus' refers to Hesperus.

(4) Jane tacitly believes that 'Venus' refers to Phosphorus.

Since claims like (1) and (2) are, strictly speaking, properly to be thought of as code for something like (3) and (4), substitution unfriendliness of tacit belief transposes into substitution unfriendliness of word reference.

This inference from cognitivism to the substitution unfriendliness of word reference is not so smooth as it seems. Even if an ascription of referential properties to words is code for a claim about some individual's idiolect, the latter constituted out of the content of their tacit beliefs, it does not follow that word reference inherits the substitution unfriendliness of tacit belief. An analogy helps to see the *non sequitur*. Suppose psychologist working on the human visual system were to assert, casually, that lines fanning out from each end of a line make that line longer, while lines arrowing out from its ends make it shorter. This would be their loose way of saying that the human visual system perceives its environment in this way. It does not follow from this that the length property of lines is as substitution unfriendly as the perception-as relation. Patently it does not: the Müller-Lyer is an illusion.

Perhaps a word's referential properties lack the objectivity of a line's geometrical properties. The length of a line is an objective feature of the world. There is more to being 15cm long than being perceived to be 15cm long. In contrast, it may be that there is nothing to a word's referring *other than* its being believed to refer. For this reason the analogy between word reference and line length just used to defeat the inference from cognitivism about word reference to the substitution unfriendliness of word reference could be unfair.

To rescue the cognitivist argument for the substitution unfriendliness of word reference in this way requires commitment either to extreme subjectivism about reference or to the believability of nonsense, even though neither is tenable. It requires both that individual speakers are capable of (tacitly) believing that a given word refers to such-and-such and either that it is meaningless to describe a word as referring *simpliciter* or that no distinction can be drawn between a word's referring and a word's being tacitly believed to refer. But belief states can be individuated from one another only if they make distinct demands on the world. If a belief's content was nonsensical, or if a belief's being held correctly amounted to no more than its being held, it would not be distinct from other beliefs with the same trait. All putative beliefs about reference would collapse into one another.[13] What the objectivity of word reference amounts to, for those who think tacit beliefs about word reference are possible and must be objective, is a further matter. It could be a threshold likelihood that the word is or can be or will be used by particular speakers to refer in some way, or it could be a convention. Whatever answer is preferred, the strengthened form of cognitivist argument against II (and hence against II$_A$ and II$_B$) seems flawed.[14]

Though it is only debatably classified as cognitivist in orientation, Jerry Fodor's account of linguistic competence provides another platform for arguing that the "puzzle" is but a phantasm, and an opportunity to review some of the counter-arguments already offered. According to Fodor 1976, Ch. 3, hearing a familiar word triggers, by default, a particular mental representation, and with luck it triggers one that is appropriate to the occasion, i.e. one with the same content as that invoked by the word's utterance. This view is often associated with the rejection of cognitivism about truth-conditional semantics for natural language, and indeed with the rejection of truth-conditional semantics for natural language *tout court*. But it does not have to be. On adopting this model of linguistic understanding one can either abandon the thought that speakers know a compositional truth theory or, alternatively, view

the model as an implementation of that very thought (see Matthews 2003 for discussion). The Fodorian model seems to offer a quick way with our puzzle no matter how it is classified. If it is a refutation of word reference (as opposed to speaker reference), including cognitivism about word reference, the puzzle disappears instantly: no word reference means no tension or inconsistency between the characteristics of word reference and the characteristic of speaker reference. If it is an implementation of cognitivism about word reference, in specifying what a word refers to we had better use words in our metalanguage that reflect the content of the mental representation associated with (i.e. habitually triggered or invoked by) the word. So if 'Venus' triggers tokenings in Jane of (her mental representation) HESPERUS rather than of (her mental representation) PHOSPHORUS, we should say that in her idiolect it refers to Hesperus, and not that it refers to Phosphorus. In that case, word reference would be substitution unfriendly.

Responses to both strategies are implicit in what has already been said. The 'refutation' line is, in effect, a form of monism. For that reason it is susceptible to the second form of the puzzle, set out in section VI. The 'implementation' line is also susceptible to that form of the puzzle (since it presupposes the representationalist theory of mind) but is, in addition, vulnerable to the criticisms of cognitivism presented in this section. A word presumably inherits its referential properties from the referential properties of the mental representation with which it is associated. Jane may not tacitly believe that 'Venus' refers to Phosphorus (even if she does tacitly believe that it refers to Hesperus), but there is no way of blocking the claim that 'Venus' *in fact* refers to Phosphorus without denying either that a word refers to whatever the mental representation it is associated with refers to or that the mental representation associated with 'Hesperus' refers to what the mental representation associated with 'Phosphorus' refers to. Neither option is attractive. By generalizing from this instance, II follows: what a word refers to, even relative to an idiolect, is substitution friendly.

VIII. Direct versus indirect theories of reference: a disjunctive solution?

I have been considering why a puzzle, or what I claim is a puzzle, has not been raised or taken seriously in this form. A fourth explanation of this phenomenon arises out of a long-standing debate over whether names are disguised definite descriptions, as Russell held, and so refer, if at all, indirectly; or whether instead they refer, if at all, directly, i.e. without the mediation of a description somehow associated with the name. Both sides in that debate face familiar and yet to be negotiated hurdles. Descriptivists need to meet the various objections set out in or adapted from Kripke 1980. Direct reference theorists, who often accept a commitment to denying substitution failure in attitude reports, need to explain away our contrary intuitions more convincingly than they have done so to date. All the same, the puzzle set out in various versions in sections III to VI, above, arguably dissolves once one accepts that at least one of the two sides is, in some form, correct.

How promising is this disjunctive stance? Arguing against it is not easy. The descriptivist/anti-descriptivist debate has spawned logically connected but self-sustaining disputes concerning attitude reports, modality, quotation, referential and attributive uses of definite descriptions, internalism versus externalism about both linguistic and mental content, information-theoretic versus conceptual-role semantics in the philosophy of mind, and so on. Inevitably this has given rise to a dense spectrum of views in place of the simple dichotomy of lore. A further question is pressing in the present context: which positions are addressing the question of how words refer (i.e. directly or indirectly), and which are addressing the question of how speakers refer? Taking each view in turn and seeing if and how it circumvents the puzzle would be an unrealistic goal in this paper. The simplest way of accommodating the profusion of theories would be for me to hand the puzzle over to you, the reader, by asking you to choose for yourself which from among the five assumptions that give

rise to the inconsistency in V you would like to reject, given your theoretical commitments in this area, and then to ask you to repeat the performance for the asterisked version of the inconsistency in VI. Rather than washing my hands entirely, though, I will persist with the pretence and offer some reasons to think that, in fact, neither of the two dominant views generates an uncomplicated solution to the puzzle.

Many direct reference theorists will look askance at I_B. Their view is associated with attempts to show, precisely, that if Lois Lane believes that Superman can fly then, appearances notwithstanding, she believes that Clark Kent can fly. What goes for believing goes for intending, with the result that if Jane, in uttering 'Venus', intends in so doing to be taken as referring to Hesperus, she also intends in so doing to be taken as referring to Phosphorus. The hope is that the contrary appearance can be explained away as a pragmatic effect.

There is something dialectically peculiar about direct reference theorists appealing to the substitution friendliness of the attitudes to see off the puzzle that is my main concern in this paper. Substitution friendliness is not a felicitous corollary of the direct reference perspective. Rather, it is a controversial lemma that needs to be argued for if one particular strategy for coping with a serious embarrassment for the perspective is to succeed. To see this, notice that unhappiness with descriptivism (for independent reasons such as those in Kripke 1980) does not make the reasons people were initially attracted to descriptivism go away. Among these reasons is the ease with which descriptivists seem able to cope with the apparent substitution unfriendliness of the attitudes. Direct reference theorists need to show either that the appearance of substitution failure is deceptive (so that descriptivist accounts of substitution failure, even if they work, are redundant), that directness of reference is in fact compatible with substitution unfriendliness (so that descriptivists are not the only ones able to account for it), or that substitution unfriendliness is a problem that arises independently of any

commitment to directness of reference (so that both direct and indirect reference theorists are on an equal footing – see e.g. Kripke 1979). The second and third strategies, for all that they succeed in the dispute between direct and indirect theorists of reference, can be pursued without rejecting I_B. Only the first strategy calls for rejection of I_B. But its rejection is a burden, not a part of the theory. Against this background, rejecting I_B because one has a direct-reference badge on one's lapel would be like appealing the possibility of causal overdetermination because one is a Cartesian dualist. Cartesians should worry about causal overdetermination, not make free appeal to it. Certainly, eye-rolling at the obviousness of the proper solution to the puzzle would be premature before the falsehood of Ib is established on a more secure independent footing than it has been.[15]

A further complication is that taking psychological attitudes to be substitution friendly with a view to making speaker reference substitution friendly may have unwelcome side-effects. In the definition of speaker reference, the name for the referent is doubly-embedded within attitude verbs. For speaker reference to be substitution friendly, both 'intends' and 'think' need to be substitution friendly. But now suppose that Lois Lane's friend, Mandy, utters 'Clark Kent is a superhero' to her. Mandy is fully aware of the identity, but she assumes that Lois is ignorant. In performing this utterance, she is in no way intent on letting Lois in on Clark's secret. She wishes merely to suggest that Lois should stop being so demanding of potential suitors, and to induce in her a belief that Clark is a superhero in, so to speak, a metrosexual kind of a way. To make sense of Mandy's utterance it is hard to avoid supposing that in uttering 'Clark Kent' she intends her action to be thought of as a reference to Clark Kent but that she does not intend her action to be thought of as a reference to Superman. Making intending and judging substitution friendly in order to make speaker reference substitution friendly sabotages this appealing description of Mandy's psychological profile.

What about indirect theories of reference? The nature of the descriptivist response to the puzzle will vary according to the nature of the descriptivism. Let us say that all forms of descriptivism treat an ordinary name, 'Venus' for example, as a definite description in disguise – 'the V', say, for some predicate 'V'. Sentences containing 'Venus' will therefore be true if, and only if, there is exactly one salient satisfier of 'V' and that satisfier meets the condition imposed by the semantics for the rest of the sentence. Strong descriptivists go further in saying, with Russell, that the definite description – and hence the name – lacks meaning in isolation. Russell's thought, here, is that, although the function of a name (and, indeed, of an explicit definite description) appears to be that of referring to an object in order that some property may then be attributed to that object by predication, the appearance is misleading. Since names do not refer, the "puzzle" of section III evaporates.

Weak descriptivists agree with strong descriptivists that names and definite descriptions fail to refer to simple entities in the world, but allow that they refer to something, for example to generalized quantifiers. Once again the puzzle disappears. 'Venus' refers, not to a planet, but to a complex abstract object. Moreover, generalized quantifiers can be discriminated as finely as the properties expressible in a description, so replacing 'V' in 'the V' with co-extensive 'V*' would change the referent. As a result, no name can have a referent that the speaker fails to appreciate is its referent, and no contradiction along the lines of the proof in section V will be possible.

While the puzzle about speaker reference and word reference gets no grip here, a simple variation on it is easy to construct, based on the notion of denotation in place of reference. An object a is denoted by a definite description 'the A', or by a name that disguises such a definite description, if and only if there is exactly one entity with the property expressed by 'A', and a is it, i.e. if and only if 'the A' refers to a generalized quantifier that determines a. Word denotation is substitution friendly. Speaker denotation seems not to be.

Suppose Jane, in uttering 'Venus', denotes Hesperus. Does it follow that she has denoted Phosphorus? It does not, if denotation is defined in the illocutionary fashion (so that what a speaker denotes is a matter of what she intends to be thought of as having denoted – see section III). Speaker denotation is a difficult notion for descriptivists, weak or strong, to do without, since they must appeal to it in accounting for the appearance of a distinction between referential and attributive uses of definite descriptions. But admitting it threatens to re-introduce the puzzle by the back door.

IX.   The puzzle solved

I have been arguing for the existence of, and explaining the failure to take seriously or even notice, a puzzle facing anyone who accepts a number of assumptions. The assumptions are that speaker reference is detemined by the speaker's intention (perhaps those of the form of I_A), that psychological attitudes permit substitution failure (I_B), that a change in the referential properties of a word alters the truth conditions of all sentences in which it appears (II_B), that truth conditions are substitution friendly (II_A), and that, normally, to refer to something it is necessary and sufficient to use a word that refers to that thing (III) – or else the asterisked versions of II_A, II_B, and III for those who are sceptical of word reference but not of mental-representation reference. The challenge has been to show that a puzzle exists. Once it is acknowledged, a solution invites itself: change I_A. More specifically, introduce an existential quantifier – an *extensionalizing existential* – into the formulation of the intention that constitutes a speaker's referring.

Let I_A' be the generalization of which the following is a particular instance:

Jane uses 'Venus' to refer to Hesperus if and only if $\exists x(x =$ Hesperus & Jane's intention in uttering the word is that her audience should think of her doing so as a reference to x).

This modification (the quantifier plus the identity-conjunct) renders the choice of metalinguistic term used to identify the referent less demanding, and so solves the puzzles instantly: speaker reference is no longer substitution unfriendly, notwithstanding intending's substitution unfriendliness. The proof of this, below, is long but straightforward – apart, perhaps, from the step to SC6. It shows how, given I_A' and the identity of Hesperus with Phosphorus, if Jane is referring to Hesperus when she uses 'Venus', she must equally be referring to Phosphorus.

P1 Jane uses 'Venus' to refer to Hesperus.

P2 Jane uses 'Venus' to refer to Hesperus iff $\exists x(x =$ Hesperus & her intention in uttering 'Venus' is that her audience should think of her doing so as a reference to x). [Instance of I_A']

SC1 $\exists x(x =$ Hesperus & Jane's intention in uttering 'Venus' is that her audience should think of her doing so as a reference to x). [From P1, P2]

P3 $c =$ Hesperus & Jane's intention in uttering 'Venus' is that her audience should think of her doing so as a reference to c. [Assumption, discharged at SC7]

SC2 $c =$ Hesperus. [From P3]

P4 Hesperus = Phosphorus.

SC3 $c =$ Phosphorus. [From SC2, P4]

SC4 Jane's intention in uttering 'Venus' is that her audience should think of her doing so as a reference to c. [From P3]

SC5 c = Phosphorus & Jane's intention in uttering 'Venus' is that her audience should think of her doing so as a reference to c. [From SC3, SC4]

SC6 $\exists x(x$ = Phosphorus & Jane's intention in uttering 'Venus' is that her audience should think of her doing so as a reference to x). [From SC5]

SC7 $\exists x(x$ = Phosphorus & Jane's intention in uttering 'Venus' is that her audience should think of her doing so as a reference to x). [From SC1, P3, and SC6 by $\exists$E]

P5   Jane uses 'Venus' to refer to Phosphorus iff $\exists x(x$ = Phosphorus & Jane's intention in uttering 'Venus' is that her audience should think of her doing so as a reference to x). [Instance of I_A']

C   Jane uses 'Venus' to refer to Phosphorus. [From SC7, P5]

SC6 uses existential generalization on a term that lies inside the scope of a psychological attitude verb. The worry is that this is no less objectionable than substituting within such a context. But the existential generalization at SC6 is immunized against the opacity of 'intends' by the identity claim in SC5, in which 'c' occurs in a non-opaque context.[16]

    This points to a different objection to I_A, sidelined since my decision at the beginning of section III to focus on just one half of the notion of intentionality. The right-hand side of I_A does not contain a commitment to the existence of the referent but its left-hand side does. A puzzle similar to the one I have been concentrating on can be mounted on the following contrast:

    $I_\exists$: Speaker reference is not existentially committed.

    $II_\exists$: Word reference is existentially committed.

That speaker reference is existentially non-committed follows from I_A and $I_{\exists B}$, a cousin of I_B.

> $I_{\exists B}$: Intending is not existentially committed (e.g. someone could intend to touch Santa Claus).

That word reference is existentially committed follows from acceptance of logical inferences such as existential generalization even in benign contexts:

> 'John is tall' is true
>
> $\therefore$ 'Something is tall' is true

This puzzle vanishes once I_A' is adopted in place of I_A. I_A', unlike I_A, requires the referent to exist. It is simply not possible for a person to refer to a non-existent. A person can still *attempt to* refer to Santa Claus, of course. To attempt to refer to Santa Claus is to attempt to bring it about that there is some entity, x, such that x is Santa Claus and one's audience recognizes one's utterance of the word 'Santa' as a reference to x. Attempting's lack of existential import makes Santa's non-existence harmless. When we say of someone, colloquially, that they referred to Santa Claus (or any other non-existent being, fictional or otherwise), we are really saying that they attempted, or perhaps feigned to attempt, to refer to Santa Claus.

I conclude, then, that accounts of speaker reference can remain intention-based without our having either to make word-reference substitution unfriendly or to give up on the intentionality of the attitudes. Showing this has not been hard. It required only the introduction of what I called an extensionalizing existential. Most of the work in this paper has been directed at showing that it needed to be shown.[17]

References

Bach, Kent (1987). *Thought and Reference*. Oxford: Clarendon Press.

Barker, Stephen J. (2004). *Renewing Meaning: A Speech-Act Theoretic Approach*. Oxford: Clarendon Press.

Bratman, Michael E. (1987). *Intention, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.

Chomsky, Noam (2000). *New Horizons in the Study of Language and Mind*. Cambridge, UK: Cambridge University Press.

Chomsky, Noam (2003). 'Reply to Egan.' In L. Antony and N. Hornstein, eds., *Chomsky and His Critics*. Oxford: Blackwell. 268-74.

Davis, Wayne A. (2005). *Nondescriptive Meaning and Reference: An Ideational Semantics*. Oxford: Clarendon Press.

Evans, Gareth (1982). *The Varieties of Reference*. John McDowell, ed. Oxford: Clarendon Press.

Fodor, Jerry A. (1976). *The Language of Thought*. Hassocks, UK: Harvester Press.

Ginet, Carl (2001). 'Deciding to Believe.' In M. Steup, ed., *Knowledge, Truth, and Duty*. Oxford: Oxford University Press.

Kripke, Saul (1976). 'Is There a Problem About Substitutional Quantification?' In G. Evans and J. McDowell, eds., *Truth and Meaning: Essays in Semantics*. Oxford: Clarendon Press. 325-419.

Kripke, Saul (1977). 'Speaker's Reference and Semantic Reference.' *Midwest Studies in Philosophy* 2: 255-76. Page references are to the reprint in G. Ostertag, *Definite Descriptions: A Reader* (Cambridge, Mass.: MIT Press).

Kripke, Saul A. (1979). 'A Puzzle about Belief.' In A. Margalit, ed., *Meaning and Use*. Dordrecht: Reidel. 239-83.

Kripke, Saul A. (1980). *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.

Larson, Richard K. and Gabriel Segal (1995). *Knowledge of Meaning: An Introduction to Semantics*. Cambridge, Mass.: MIT Press.

Laurence, Stephen (1996). 'A Chomskian Alternative to Convention-Based Semantics.' *Mind* 105(418): 269-301.

Linsky, L. (1963). 'Reference and referents.' In C. E. Caton, ed., *Philosophy and Ordinary Language*. Urbana: University of Illinois Press. 74-89.

Matthews, Robert J. (2003). 'Does Linguistic Competence Require Knowledge of Language?' In A. Barber, ed., *Epistemology of Language*. Oxford: Oxford University Press. 187-213.

McGilvray, James (1998). 'Meanings Are Syntactically Individuated and Found in the Head.' *Mind and Language* 13(2): 225 80.

McGilvray, James (2005). *The Cambridge Companion to Chomsky*. Cambridge: Cambridge University Press.

Mellor, W. W. (1954). 'Believing the Meaningless.' *Analysis* 15: 41-43.

Mercier, Adele (2003). 'Are Language Conventions Philosophically Explanatory? (Or: "It's Shirt-Buttoning All the Way Down, Ruth!").' *Croatian Journal of Philosophy* 3(8): 111-24.

Pietroski, Paul M. (2003). 'The Character of Natural Language Semantics.' In A. Barber, ed., *Epistemology of Language*. Oxford: Oxford University Press. 217-56.

Quine, W. V. (1990). *Pursuit of Truth*. Cambridge, Mass.: Harvard University Press.

Recanati, Francois (1997). 'Can We Believe What We Do Not Understand?' *Mind and Language* 12(1): 84-100.

Salmon, Nathan (1986). *Frege's Puzzle*. Cambridge, Mass.: MIT Press.

Salmon, Nathan (2004). 'The Good, the Bad, and the Ugly.' In M. Reimer and A. Bezuidenhout, eds., *Descriptions and Beyond*. Oxford: Clarendon Press. 230-60.

Saul, Jennifer M. (1996). 'What's Wrong with Metalinguistic Views.' *Acta Analytica* 16/17: 81-94.

Searle, John R. (1969). *Speech Acts: an Essay in the Philosophy of Language*. London: Cambridge University Press.

Strawson, P. F. (1950). 'On Referring.' *Mind* 59: 320-44.

Stroll, Avrum (1953). 'A Problem Concerning the Analysis of Belief Sentences.' *Analysis* 14: 15-19.

Stroll, Avrum (1955). 'Believing the Meaningless; a Reply to Mr Mellor.' *Analysis*.

Swift, Jonathan, ed. (1726). *A voyage to Laputa, Balnibarbi, Glubbdubdrib, Luggnag and Japan*. Book III of Volume 2 of *Travels into Several Remote Nations of the World. By Captain Lemeul Gulliver*. London: Benjamin Motte.

---

[1] Davis 2005 seems to take seriously what it is 'natural' to say about what a person has referred to; and Chomsky

2003: 272, citing Strawson, writes that whereas talk of speakers referring is 'normal in English (with

counterparts in other languages)', talk of words referring is 'a technical innovation that remains obscure and unexplained', and infers from this that '[t]hose who believe that some notion of "word-thing reference"' has an explanatory role 'have to explain what they mean, and provide the evidence'.

[2] Thus, Kripke's distinction is designed to call into question Donnellan's criticism of Russell's account of definite descriptions, while Quine's claim is driven, presumably, by a concern with logical relations whose study has traditionally involved the interpretation of the elements of formal languages.

[3] Compare Strawson 1950: 333. One might accept this point but still see in Russell's theory the seeds of a useful explanation of our capacity to *think* about non-existent objects. In that case, his theory of descriptions would feed into a theory of mental content, not linguistic content, and be relevant to attempts to refer or to avoid that fail through the lack of an object.

[4] I will follow a policy throughout of illustrating I_A, etc., using instances rather than expressing them as generalizations. Generalizations would need to be expressed schematically rather than using objectual quantification into the complement clause of psychological operators, and using schematic generalization does not illuminate anything.

[5] The reasons for this are similar to the reasons for why beliefs must not be true merely because they are held – something I argue for explicitly in section VII.

[6] The version of the proof below, thought still informal, is detailed enough to reveal how the contribution of IIa distributes out into uncontroversial uses of Hypothetical Syllogism and the Law of Identity (see main text). Moreover, the contribution of IIb dissolves in to the re-use of the background semantics for the rest of the sentence (i.e. P2, as used in going from SC3 to Conc.).

>   P1. 'Venus' refers Hesperus.
>
>   P2. [Semantics for 'Venus is wet' minus P1]
>
>   SC1. 'Venus is wet' is true iff Hesperus is wet. (From P1, P2)
>
>   P3. Hesperus is Phosphorus.
>
>   SC2. Hesperus is wet iff Phosphorus is wet. (From P3 using the Law of Identity uncontroversially.)
>
>   SC3. 'Venus is wet' is true iff Phosphorus is wet. (From SC1 & SC2, in an uncontroversial use of HS)
>
>   Conc. 'Venus' refers to Phosphorus. (From SC3, P2)

---

[7] The fallback position I am urging against here is more clearly problematic in the version of the puzzle that figures in section VI. It is difficult to see how a speaker could use a word referringly without using it to refer to what the mental representation they thereby invoke refers. (See the next section for a definition of 'invoke'.)

[8] Whether the relation is in fact four-place – a speaker$_I$ using a word$_{II}$ to refer some audience$_{III}$ to an object$_{IV}$ – is not a pressing issue here. My own view is that the actual existence of an audience is unnecessary. What matters is that the speaker thinks there is an audience, since only then can he or she form the intention to have that audience think of the act as a reference to the object. A similar line cannot be taken on the existence of the object. One cannot refer save to something that really exists. I touch on this in section IX.

[9] There is a question over the nature of the shorthand. It is hardly synecdoche or metonymy. 'The (hired) gun killed the child' in a Raymond Chandler novel means that the person carrying the gun killed the child; but while the NRA slogan may be confused, it is not so confused as to be saying that people don't kill people, people kill people. The shorthand is more plausibly seen as a form of ellipsis. The criticism I offer of the 'shorthand' position is neutral on how it is developed.

[10] One line to take with a voluntarist about belief is to challenge them to form the belief that there is a human-sized pigeon behind them; if they claim to have met the challenge, bet them €1,000 that there is no such pigeon. By refusing they will be implicitly conceding the argument. Were they to accept, you would at least end up a little wealthier. In an evolutionary context, the latter outcome would be equivalent to them being at a reproductive disadvantage.

[11] For more on constraints on intention formation, see Bratman 1987; for more on doxastic anti-voluntarism see Ginet 2001.

[12] It is not clear, for example, whether the anti-monist argument requires an information-theoretic theory of content of the kind criticised by McGilvray (echoing Chomsky) in the quotation in section I rather than, say, a more interpretationist semantics such as that apparently endorsed in Chomsky 2000: 160, and more familiarly associated with Daniel Dennett.

[13] Similarly, states of intending are distinct only if acting on them successfully requires distinct outcomes. This is why referring cannot be a matter of intending to refer to it. See section III and the discussion of Ib.

[14] For an explicit if slightly dated exchange on the believability of nonsense, see the exchange between Avrum Stroll 1953; 1955 and W. W. Mellor 1954. François Recanati 1997 allows that nonsense can be believed, but predicates this possibility on the believer hearing and deferring to a speaker without realizing that the words

coming out of the latter's mouth have no content. Whatever its merits, applying this explanation in the present context would mean treating all tacit belief about reference as parasitic on explicit talk about reference. This would fail to explain how people sheltered from such explicit talk may nevertheless be linguistically competent.

[15] As it happens, the first strategy, when it has been followed, has led to a theory that resembles the solution offered to the puzzle in section IX, even though this latter does not involve rejecting I_B. According to Salmon 1986, the belief relation, which is a two place relation, is derivative of a three-place relation, BEL(…), which holds between a person, a Russellian proposition, and one of that proposition's guises. *S* believes that *p* if and only if $\exists x$(BEL(*S*, *p*, *x*)). *Believes* is therefore insensitive to the choice among distinct but co-referring terms to pick out its relata, but the choice of word pragmatically implicates a particular guise. Whether this use of (what I will call) an extensionalizing existential succeeds is a moot point, but it is possible to use a similar sounding strategy to make speaker reference substitution friendly without rejecting I_B (and without quantifying over guises).

[16] In earlier versions of the paper I experimented with using substitutional quantification to avoid the worry. I now think that on balance an objectual treatment is adequate, and avoids the need to confront anxieties about both substitutional quantification *per se* (see Kripke 1976) and 'metalinguistic' approaches to opacity (see Saul 1996). Worth noting is that Searle 1969 includes a substitutionally interpreted existential quantifier in his 'rules of reference' (pp. 94-6; the interpretation of the quantifier is discussed on p. 94). This seems designed to block existential failure (to be discussed below in the main text) rather than substitutional failure, as his rules lack anything equivalent to the 'x = Hesperus', and remarks on pp. 89-90 indicate reluctance to include one. Indeed, he steers around the puzzle by defining (the equivalent of) 'Jane refers' rather than 'Jane refers to Hesperus'. In the preliminaries to the rules, he assimilates puzzles about substitution to what I earlier called II' (pp. 79, 90-1) and an 'axiom of identification' (p. 80). Of course, his account could be adapted. My emphasis has been on failure to spot the puzzle, not on the difficulty of making changes to cope with it once it is spotted.

[17] This paper has evolved out of talks with the same title at (in chronological order) the *Proper Names* workshop at the Humboldt Institute in Berlin, the philosophy departments of Sussex University, Birmingham University, and the Open University, the *Pragmatics* conference in Riva del Garda, the *Canadian Philosophical Association* at the University of Western Ontario, the *Joint Session of the Aristotelian Society and Mind Association* in Southampton, and the *Philosophy of Linguistics* workshop at the Inter-University Centre in Dubrovnik. My