White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/132147/

Version: Published Version

Article:
Trueman, Robert (2018) Substitution in a sense. Philosophical Studies. pp. 3069-3098. ISSN: 0031-8116

https://doi.org/10.1007/s11098-017-0995-1

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Philos Stud
https://doi.org/10.1007/s11098-017-0995-1

Substitution in a sense

Robert Trueman[1]

© The Author(s) 2017. This article is an open access publication

Abstract The Reference Principle (RP) states that co-referring expressions are everywhere intersubstitutable salva congruitate. On first glance, (RP) looks like a truism, but a truism with some bite: (RP) transforms difficult philosophical questions about co-reference into easy grammatical questions about substitutability. This has led a number of philosophers to think that we can use (RP) to make short work of certain longstanding metaphysical debates. For example, it has been suggested that all we need to do to show that the predicate '( ) is a horse' does not refer to a property is point out that '( ) is a horse' and 'the property of being a horse' are not everywhere intersubstitutable salva congruitate. However, when we understand 'substitution' in the simplest and most straightforward way, (RP) is no truism; in fact, natural languages are full of counterexamples to the principle. In this paper, I introduce a new notion of substitution, and then develop and argue for a version of (RP) that is immune to these counterexamples. Along the way I touch on the following topics: the relation between argument forms and their natural language instances; the reification of sense; the difference between terms and predicates; and the relation between reference and disquotation. I end by arguing that my new version of (RP) cannot be used to settle metaphysical debates quite as easily as some philosophers would like.

Keywords Substitution · The Reference Principle · Sense · Reference · Disquotation · Singular terms · Predicates · Properties · Propositions

✉ Robert Trueman
rob.trueman@york.ac.uk

[1] Department of Philosophy, University of York, York, United Kingdom

Published online: 25 October 2017

1 Introduction

It is always tempting to look for shortcuts in philosophy, for easy answers to difficult questions. Here are two examples of a shortcut that some philosophers have tried:

> **Easy Argument 1: Properties**
> Do predicates refer to properties? Well, if the predicate '( ) is a horse' refers to a property, then it must surely refer to the property of being a horse. And in that case, '( ) is a horse' and 'the property of being a horse' must co-refer. But co-referring expressions are everywhere intersubstitutable *salva congruitate*. (In other words, if two expressions co-refer then substituting the one for the other can never turn a grammatical sentence into an ungrammatical one.) '( ) is a horse' and 'the property of being a horse' are not everywhere intersubstitutable *salva congruitate*: 'Shergar is a horse' is a grammatical sentence, but 'Shergar the property of being a horse' is not. So '( ) is a horse' does not co-refer with 'the property of being a horse', and thus '( ) is a horse' does not refer to a property.[1]

> **Easy Argument 2: Propositions**
> Are propositional attitudes relations to propositions? Take the sentence 'Sharon hopes that the show will be funny'. If this sentence says that Sharon bears the relation of hoping to the proposition that the show will be funny, then 'that the show will be funny' must surely refer to that proposition. And in that case, 'that the show will be funny' must co-refer with 'the proposition that the show will be funny'. But again, these expressions are not everywhere intersubstitutable *salva congruitate*: 'Sharon hopes that the show will be funny' is a grammatical sentence, but 'Sharon hopes the proposition that the show will be funny' is not. So 'that the show will be funny' does not refer to a proposition in 'Sharon hopes that the show will be funny', and thus the propositional attitude of hoping, at least, is not a relation to a proposition.[2]

These easy arguments crucially rely upon the *Reference Principle*:

> (RP) Co-referring expressions are everywhere intersubstitutable *salva congruitate*.[3]

Of course, that is not all that these arguments rely upon. For example, Easy Argument 1 presumes that if '( ) is a horse' refers to a property, then that property

[1] This argument is given by Wright in his (1998: pp. 73, 85–90); however, Wright (pp. 86–88) tempers its conclusion by insisting that although predicates do not refer to properties, they do 'ascribe' them. Moltmann offers a related argument in her (2003a: Sect. 3).

[2] For this and related arguments, see: Bach (1997: p. 225), McKinsey (1999: pp. 530–531), Moltmann (2003b: Sect. 2.1), Rosefeldt (2008: Sect. 3).

[3] The name 'the Reference Principle' is due to Wright (1998: p. 73). In fact, my Reference Principle is just one half of Wright's; however, it is the only half that is needed for the arguments presented above. It is also worth mentioning that Hale and Wright use 'the Reference Principle' for something quite different in their (2012: p. 93).

can also be referred to by a singular term, and that presumption can be challenged.[4] Nonetheless, (RP) is clearly the driving force behind these easy arguments; it transforms difficult philosophical questions about co-reference into easy grammatical questions about intersubstitutability. What is more, (RP) initially appears to be a truism. Indeed, at first we really want to demand something stronger: co-referring expressions should not just be everywhere intersubstitutable *salva congruitate*, but *salva veritate* as well! Most of us eventually abandon this stronger demand when we are confronted with the stubborn intensionality of natural language, but still, intensionality does not seem to do anything to threaten the minimal *salva congruitate* requirement.

So it seems that (RP) is an uncontentious principle which makes quick work of knotty philosophical questions. This all sounds too good to be true. And of course, it is. In his (2005), Oliver offers a number of examples designed to show that far from being a truism, (RP) is demonstrably false.[5] For now it will suffice to focus on just two of Oliver's examples. First, take the sentence,

(1a) Brilliant Bertrand solved a paradox.

'Bertrand' and 'the referent of "Bertrand" ' co-refer,[6] but the result of substituting 'the referent of "Bertrand" ' for 'Bertrand' in (1a),

(1b) Brilliant the referent of 'Bertrand' solved a paradox,

is an ungrammatical muddle of words. Second, take the sentence,

(2a) I am Robert.

'I' and 'me' co-refer in any given context, but again, the result of substituting 'me' for 'I' in (2a),

(2b) Me am Robert,

is at very least not the Queen's English.[7]

What are we to make of these counterexamples to (RP)? Well, they both rely on a particular conception of substitution, which I will call *simple-substitution*. According to this conception, we substitute one expression for another by (almost literally) cutting out the latter and pasting the former in its place. (This is the sort of substitution that you can use the Find and Replace function on a word processor to do.) Oliver has therefore presented counterexamples to the following precisified version of (RP):

[4] In fact, it should be challenged. See Trueman (2015).

[5] Similar examples are given by Schiffer (2003: pp. 92–95).

[6] Or at least they do on the assumption that 'Bertrand' and 'the referent of "Bertrand" ' are referring singular terms. As MacBride (2011: pp. 302–304) points out, this assumption can be challenged. However, for the purposes of this paper, I will happily concede to Oliver that 'Bertrand' and 'the referent of "Bertrand" ' are referring singular terms.

[7] Rosefeldt's (2008: p. 309) supposedly fixed version of (RP) is not immune to this counterexample.

(RP_1) Co-referring expressions are everywhere simple-intersubstitutable *salva congruitate*.

Now, we cannot criticise Oliver for being uncharitable when he interprets (RP) as (RP_1): the conception of substitution involved in the easy arguments certainly seems to be simple-substitution. Nonetheless, we might wonder whether this really is the best way to interpret (RP). While simple-substitution may be appropriate for well designed formal languages, it is surely *too* simple when it comes to messy natural languages. My aim in this paper is to present an alternative conception of substitution. I will then develop and argue for a version of (RP) that is immune to Oliver's counterexamples. Finally, at the very end of the paper, I will return to the easy arguments with which we began; as we will see, when we rest these arguments on my new version of (RP), they become a lot less easy.

2 Sense-substitution

Consider the following argument:

(i) Robert is hungry
(ii) I am Robert
$\therefore$ (iii) I am hungry

This argument is obviously valid. Indeed, I would say that it is a natural language instance of the following valid argument form:

(SI) $Fa;\ b = a;\ \therefore Fb.$

However, the validity of (SI) crucially turns on our substituting '*b*' for '*a*' in '*Fa*'. So if (i)–(iii) is to be an instance of (SI), then (iii) must in some sense count as a result of substituting 'I' for 'Robert' in (i). And clearly, we do not have a case of simple-substitution here. As we move from (i) via (ii) to (iii), we do not unthinkingly write 'I' into the gap in '( ) is hungry'. Rather, we swap '( ) is hungry' for '( ) am hungry', and then write 'I' into that gap instead. Moreover, in making this swap we in no way impugn the validity of the argument. As far as the argument is concerned, (iii) really is a result of substituting 'I' for 'Robert' in (i). What we must now ask, then, is: Why should (iii) count as a result of this substitution?

I would like to propose the following answer: Because '( ) am hungry' and '( ) is hungry' have the same *sense*, in something like Frege's sense of 'sense'. As Frege thought of it, the sense of a sentence (which I will also call a *proposition*) determines the logical properties of that sentence,[8] and the sense of a subsentential expression is that expression's contribution to the senses of the sentences in which it

[8] In fact, Frege (e.g. 1906: pp. 318, 332) made the stronger claim that logical relations hold primarily between propositions, and only derivatively between sentences; for Frege (1918: pp. 334–347), propositions are abstract objects, and it is these abstract objects that are the primary relata of logical relations. I, however, would like to distance myself from the whole idea that senses are a kind of object (see Sect. 4.1).

appears.[9] So the idea here is that the grammatical difference between '( ) am hungry' and '( ) is hungry' does not reflect a difference in the contributions that these expressions make to the senses of the sentences in which they appear. If that is right, then we can explain why (i)–(iii) should count as an instance of (SI): '( ) is hungry' and '( ) am hungry' make exactly the same contribution to the senses of (i) and (iii), and it is those senses that determine the logical properties of those sentences.

More certainly needs to be said about the notion of sense appealed to in this answer, and I will try to say more in Sect. 4. For now, though, I would ask the reader to go with my answer, if only to see where it leads: (iii) counts as a result of substituting 'I' for 'Robert' in (i) because '( ) am hungry' has the same sense as '( ) is hungry'.[10] This suggests that there is a conception of substitution according to which the following is true:

> $\Psi(\beta)$ is a result of substituting β for the displayed occurrence of α in $\Phi(\alpha)$ whenever $\Psi(\ ) \approx \Phi(\ )$,[11]

where '$\approx$' is shorthand for 'has the same sense as', and 'the displayed occurrence of α in $\Phi(\alpha)$' is shorthand for 'the occurrence of α in $\Phi(\alpha)$ that fills the gap in $\Phi(\ )$'. The 'displayed occurrence' clause is strictly necessary, since there may be other occurrences of α in $\Phi(\alpha)$, but for the sake of readability I will often leave it implicit.

Now consider this argument:

(iv) Tim helped Robert with a paper
(v) I am Robert
∴ (vi) Tim helped me with a paper

Again, this argument is obviously valid, and again I would say that it is a natural language instance of (SI). But as we move from (iv) via (v) to (vi), we do not unthinkingly write 'I' into the gap in 'Tim helped ( ) with a paper'. This time, we swap 'I' for 'me' and write that into the gap instead. Moreover, in making this swap we in no way impugn the validity of the argument. As far as the argument is concerned, (vi) really is a result of substituting 'I' for 'Robert' in (iv). And now, of course, what we must ask is: Why should (vi) count as a result of this substitution?

I would like to propose the following answer: Because 'I' and 'me' have the same sense (in any given context, at least).[12] Again, the idea is that the grammatical difference between 'I' and 'me' does not reflect a difference in the contributions that

[9] See (Frege 1893: Sect. 32). Here and elsewhere (e.g. 1923: p. 390), Frege uses a mereological metaphor: he says that the sense of a part of a sentence is a part of the sense of that sentence. I do not find that metaphor very helpful, not least because it encourages us to think of senses as objects.

[10] I am by no means suggesting that this is the only possible answer. I discuss an alternative in the "Appendix".

[11] If we were being fastidious, we would put Quine-quotes around '$\Psi(\beta)$' and '$\Phi(\alpha)$'. However, Quine-quotes tend to make things harder to understand, and so I will only use them when I really have to. For the rest of the time, we should just imagine that Quine-quotes are already built into the Greek variables.

[12] From now on, I will leave the 'in any given context' qualification tacit.

these expressions make to the senses of the sentences in which they appear.[13] If that is right, then we can explain why (iv)–(vi) should count as an instance of (SI): 'I' and 'me' make exactly the same contribution to the senses of the sentences in which they appear, and so for logical purposes we are free to swap between them as grammar demands. This suggests that there is a conception of substitution according to which the following is true:

> $\Phi(\gamma)$ is a result of substituting β for the displayed occurrence of α in $\Phi(\alpha)$ whenever $\gamma \approx \beta$.

Putting these answers together, I would like to introduce the following conception of substitution, which I will call *sense-substitution*:

> To sense-substitute β for the displayed occurrence of α in a meaningful sentence $\Phi(\alpha)$ is to construct a meaningful sentence $\Psi(\gamma)$ such that:
>
> (a) $\Psi(\ ) \approx \Phi(\ )$
> and (b) $\gamma \approx \beta$.

We will work through some examples of sense-substitution in the next section, but for now we can explain the guiding idea like this. Simple-substitution is appropriate when we are interested in the straightforward substitution of expressions thought of as ink marks on a page (or lights on a computer screen, or whatever). Sense-substitution, on the other hand, is the conception to use when we are interested in the substitution of *senses*, not ink marks. Of course, we cannot directly substitute one sense for another; all we can really do is substitute one expression for another. But when we are interested in the senses of these expressions, we can simply disregard what they look like. Whether $\Psi(\gamma)$ counts as a result of sense-substituting β for α in $\Phi(\alpha)$ has nothing to do with typography. $\Psi(\ )$ and $\Phi(\ )$ do not need to share any typographic similarity, and neither do γ and β. All that matters is that $\Psi(\ )$ have the same sense as $\Phi(\ )$, and γ have the same sense as β.

[13] Of course, this does not mean that differences in case are pointless grammatical flourishes. Like word order, case helps to determine which terms fill which argument places in a given predicate. Or to use the language of thematic roles: case helps to determine which term refers to the Agent of the relation expressed by the given predicate, which to the Patient, which to the Instrument, and so on. Indeed, it seems to me that the suggestion that differences in case do not reflect differences in sense can be made to fit very well with a thematic-role analysis. On one such analysis, the sentences (a) 'I phoned Sharon' and (b) 'Sharon phoned me' become (a′) '$\exists e(\text{phoning}(e) \wedge \text{Agent}(i, e) \wedge \text{Patient}(s, e))$' and (b′) '$\exists e(\text{phoning}(e) \wedge \text{Agent}(s, e) \wedge \text{Patient}(i, e))$'. The idea that 'I' and 'me' have the same sense in (a) and (b) is borne out here by the fact that both of these terms are formalised as 'i' in (a′) and (b′); the difference in case simply tells us whether 'i' appears in 'Agent$(\ , e)$' or in 'Patient$(\ , e)$'. For an introduction to thematic-role analysis, and the hypothesised link between thematic roles and grammatical roles (often called UTAH), see: Larson and Segal (1995: Sect. 12.3), Heim and Kratzer (1998: pp. 49–58).

3 A new Reference Principle

We can use the notion of sense-substitution to give a new precisification of (RP):

(RP_2) Co-referring expressions are everywhere sense-intersubstitutable.

Conspicuously, I have not included a '*salva congruitate*' clause in (RP_2). This is because sense-substitution is defined only over meaningful sentences: in the definition of sense-substitution, I stipulated that $\Phi(\alpha)$ and $\Psi(\gamma)$ be *meaningful* sentences. I made this stipulation because sense-substitution is meant to give us a way of substituting one sense in a proposition for another, and only meaningful sentences express propositions. (RP_2) should therefore be understood as follows: if α and β co-refer, then there is no occurrence of α in a *meaningful* sentence such that it is impossible to construct another *meaningful* sentence that counts as a result of sense-substituting β for that occurrence of α. Read in this way, there is no need to include an extra '*salva congruitate*' clause.

(RP_2) can be defended from Oliver's counterexamples. Let us again start with,

(1a) Brilliant Bertrand solved a paradox.

According to (RP_2), since 'Bertrand' co-refers with 'the referent of "Bertrand" ', it should be possible to sense-substitute the latter for the former in (1a). My suggestion is that the following counts as a result of this sense-substitution:

(1c) The referent of 'Bertrand' is brilliant and solved a paradox.

(1c) is a result of this sense-substitution if:

(a) '( ) is brilliant and solved a paradox' ≈ 'Brilliant ( ) solved a paradox'

and (b) 'the referent of "Bertrand" ' ≈ 'the referent of "Bertrand" '.

(a) strikes me as plausible,[14] and we can take (b) for granted; more generally, we can take it for granted that every expression has the same sense as itself.[15] However, it is worth noting that I am suggesting only that (1c) is *a*, not *the*, result of this sense-substitution. Unlike simple-substitution, more than one sentence can count as a result of sense-substituting one expression for another. Take the following sentence for example:

(1d) The thing to which 'Bertrand' refers is brilliant and solved a paradox.

Presumably, 'the thing to which "Bertrand" refers' has the same sense as 'the referent of "Bertrand" '. If so, (1c) is a result of sense-substituting 'the referent of "Bertrand" ' for 'Bertrand' in (1a) if and only if (1d) is as well.

[14] On one way of reading (1a). Alternatively, we might read (1a) as saying of brilliant Bertrand that he solved a paradox. On this alternative reading, (1a) is equivalent to 'The thing that is brilliant and identical to Bertrand solved a paradox'. If we read (1a) in this way, then we should obviously give the following instead of (1c): The thing that is brilliant and identical to the referent of 'Bertrand' solved a paradox. Or more colloquially: The brilliant referent of 'Bertrand' solved a paradox.

[15] More generally still, I will assume that *has the same sense as* is an equivalence relation over meaningful expressions. This is obviously a simplifying idealisation; see Sect. 4.3 for a very brief discussion.

We turn now to,

(2a) I am Robert.

According to (RP_2), since 'I' co-refers with 'me', it should be possible to sense-substitute the latter for the former in (2a). I would like to suggest that (2a) is *itself* a result of sense-substituting 'me' for 'I' in (2a). (2a) counts as a result of this sense-substitution if:

(a) '( ) am Robert' ≈ '( ) am Robert'
and (b) 'I' ≈ 'me'.

In this case it is (a) that we can take for granted, and (b) that is plausible. So, it seems, (2a) is a result of sense-substituting 'me' for 'I' in (2a). In other words, substituting 'me' for 'I' is a limiting case of sense-substitution, like substituting 'I' for 'I' or 'Bertrand' for 'Bertrand'.[16] Again, I am claiming only that (2a) is *a* result of sense-substituting 'me' for 'I' in (2a). In fact, there is no reason to limit our attention to English sentences when looking for results of this sense-substitution. Consider the French sentence,

(2c) Je suis Robert.

This sentence is a result of sense-substituting 'me' for 'I' in (2a) if:

(a) '( ) suis Robert' ≈ '( ) am Robert'
and (b) 'Je' ≈ 'me'.

(a) looks like a safe assumption; certainly, '( ) suis Robert' is the French translation of '( ) am Robert', and while translation may not always preserve sense, it is hard to see what the difference could be in this case. As for (b), 'Je' is the French translation of 'I', and again it seems reasonable to say that in this case, translation preserves sense; if we add to this the earlier assumption that 'I' and 'me' also have the same sense, we can infer (b).[17] In short, then, (2a) is a result of sense-substituting 'me' for 'I' in (2a) if and only if (2c) is as well.

This brings us to an important point. When we ask whether one expression can be sense-substituted for another, the modality we have in mind should be a very permissive one. We should be asking whether it is *in principle possible* for there to be a sentence which counts as a result of this sense-substitution. After all, sense-substitution is meant to give us a way of substituting senses in propositions. So, saying that β can be sense-substituted for α in $\Phi(\alpha)$ is meant to capture the idea that we can substitute the sense of β for the sense of α in the proposition expressed by $\Phi(\alpha)$. It should not matter, then, if the proposition that results from this substitution of senses cannot be expressed in the same language as $\Phi(\alpha)$. It should not even matter if it cannot be expressed in any of the languages that anyone has ever actually spoken. All that matters is that it is *in principle possible* to construct a sentence that

[16] I first got the idea that substituting 'me' for 'I' might be a limiting case of substitution from Dolby (2009: p. 291).

[17] I am again idealising by assuming that *has the same sense as* is an equivalence relation, and hence transitive, over meaningful expressions; again, see Sect. 4.3.

expresses the proposition, even if doing so involves inventing a whole new language. Of course, it is a very nice question exactly how we should characterise this permissive modality, but I hope that I have said enough for present purposes. In what follows, then, I will understand the modality involved in 'sense-intersubstitutable' in this permissive sort of way.

This point is related to a counterexample to (RP_1) that we have not yet considered. 'Frege' co-refers with 'Der Autor der *Grundlagen der Arithmetik*', and so (RP_1) entails that they are everywhere simple-intersubstitutable *salva congruitate*. However, whereas,

(3a) Frege was a philosopher,

is a grammatical English sentence,

(3b) Der Autor der *Grundlagen der Arithmetik* was a philosopher,

is not a grammatical sentence in English or German. This is another of Oliver's (2005: p. 186) counterexamples, but he does not put too much weight on it. As he concedes, it can be disarmed by restricting (RP_1) so that it applies to only one language at a time; what the counterexample shows is just that (RP_1) should not be taken to imply that co-referring expressions *taken from different languages* are everywhere intersubstitutable *salva congruitate*. However, this clash of languages does nothing to stop us *sense*-substituting 'Der Autor der *Grundlagen der Arithmetik*' for 'Frege' in (3a). 'The author of *The Foundations of Arithmetic*' presumably has the same sense as 'Der Autor der *Grundlagen der Arithmetik*', and so one result of the sense-substitution will be:

(3c) The author of *The Foundations of Arithmetic* was a philosopher.

And assuming that the German predicate '( ) war ein Philosoph' has the same sense as the English predicate '( ) was a philosopher', another result will be:

(3d) Der Autor der *Grundlagen der Arithmetik* war ein Philosoph.

And of course, there will be plenty of other results of this substitution, not just in German and English, but in innumerable other languages as well. (RP_2) need not, then, be restricted to one language at a time. More than that, it should not be: again, to say that β can be sense-substituted for α in $\Phi(\alpha)$ is meant to capture the idea that we can substitute the sense of β for the sense of α in the proposition expressed by $\Phi(\alpha)$, and so it really should not matter whether α and β belong to the same language. I will, then, take (RP_2) to apply simultaneously to *all* languages, including merely possible ones.

In summary, (RP_2) is not vulnerable to any of Oliver's counterexamples to (RP_1). Or at least, not *obviously* vulnerable.[18] It is impossible to miss the slightly hesitant tone running throughout the discussion so far. I have made lots of 'safe

[18] In fact, even this is a stronger claim than I have any right to make. Oliver offers a dizzying array of counterexamples to (RP_1), and I have only presented those that I find most compelling. Of course, it goes without saying that I think that (RP_2) can also be defended from the rest of Oliver's counterexamples, but unfortunately, I do not have the space to discuss them here.

assumptions' and 'plausible suggestions' about whether one expression has the same sense as another. And really, I cannot be more definitive than that. It is notoriously difficult to show once and for all that two expressions have exactly the same sense. The difficulties here are in part philosophical—just how finely grained should our concept of sense be?—but also in part empirical: demonstrating that two expressions share a sense will certainly require a close examination of how those expressions are actually used. It is, then, far beyond the means of this paper to announce once and for all that (1c) is a result of sense-substituting 'the referent of "Bertrand" ' for 'Bertrand' in (1a), or that (2a) is itself a result of sense-substituting 'me' for 'I' in (2a), and so on through the other examples. Nonetheless, what I have said is enough to undermine Oliver's counterexamples, taken as counterexamples to (RP_2) rather than (RP_1). As of yet we have been given no reason to suspect that there might be co-referring expressions that are not everywhere sense-intersubstitutable.

Still, it would be nice not to end on such an uncertain note. I will, therefore, spend most of what remains of this paper working towards a general argument for a version of (RP). But before starting on that work, I would like to deal with a different objection to (RP_2), namely the objection that there is something suspect about the notion of sense in sense-substitution.

4 The sense in sense-substitution

There are two different kinds of worry that you might have about my appeal to sense: first, you might have a general worry about the whole idea of sense, and second, you might have a particular worry about the details of *my* appeal to sense. In this section I will try to answer both kinds of worry.

4.1 The general worry: the very idea of sense

In Sect. 2 I made it clear that I was working with a broadly Fregean notion of sense, and I am sure that that will have set alarm bells ringing for some readers. But I think that we can address this sort of worry simply by emphasising the 'broadly' in 'broadly Fregean'. I have so far made only two substantial assumptions about sense:[19] first, that the sense of a sentence determines the logical properties of that sentence, and second, that the sense of a subsentential expression is that expression's contribution to the senses of the sentences in which it appears. And while it is true that I took both of these assumptions from Frege, they are hardly peculiarly Fregean. Accepting them certainly does not force us to accept any of Frege's more controversial claims about sense. This point is well illustrated by the following example. One of Frege's most famous doctrines is that (even simple) singular terms can co-refer and yet have different senses (1892: pp. 151–152). Now,

[19] So far? In Sect. 6 I will make some further claims about the senses of singular terms, first-level predicates and second-level predicates. However, I think that these further claims should be seen as substantial assumptions about terms and predicates, not about sense.

I am inclined to agree with Frege on this issue, but for all I have said in this paper, it may be that co-referring (simple) singular terms automatically share a sense, i.e. automatically make the same contribution to the senses of the sentences in which they appear.

I hope that this already goes some way towards easing the general worry that one might have about my appeal to sense, but I would like to go further. Frege (1892: p. 154, 1918: pp. 334–337) thought that senses were *objects*; that is, Frege thought that senses could be referred to with singular terms. However, some philosophers (e.g. McDowell 1977) think that it is a mistake to reify senses in this way. This is not to deny that one expression can have the same sense as another. It is just to deny that what it is for α to have the same sense as β is for α and β to stand in the *having* relation to the same special object, which we call their 'sense'. Instead, to say that α has the same sense as β is just a flowery way of saying that α is synonymous with β (for logical purposes at least). It is important to observe, then, that my definition of sense-substitution does not treat senses as if they were objects: sense is only mentioned in the course of saying that one expression *has the same sense as* another.[20]

At this point it might be objected that while my definition of sense-substitution does not treat senses as objects, my less formal remarks about sense do. In fact, my two assumptions about sense are cases in point: I assumed that *the sense* of a sentence determines the logical properties of that sentence, and that *the sense* of a subsentential expression is that expression's contribution to *the senses* of the sentences in which it appears. However, it is my belief that we can offer explanations—or more accurately: *explications*—of everything I say about sense that do not treat senses as objects. Again, take my two assumptions about sense as examples. It is relatively easy to offer an explication of the claim that the sense of a sentence determines that sentence's logical properties: if two sentences have the same sense, then they have the same logical properties.[21] It is a little harder to offer an explication of the claim that the sense of a subsentential expression is its contribution to the senses of the sentences in which it appears, but here is a start: subsentential expressions α and β have the same sense if and only if substituting β for α in a sentence always produces a sentence with the same sense as the original, and *vice versa*.[22]

[20] The one point in the definition at which this is not *absolutely* explicit is when I stipulate that $\Phi(\alpha)$ and $\Psi(\gamma)$ are *meaningful* sentences; however, we can understand the claim that a sentence is meaningful as the claim that it has the same sense as itself.

[21] Importantly, this does not automatically imply that two sentences have the same sense just in case they entail each other. To begin with, all that is being claimed here is that sentences with the same sense have the same logical properties, not that sentences with the same logical properties also have the same sense. And perhaps more interestingly, even if we did want to say that sentences with the same logical properties have the same sense, we could still insist that sentences have logical properties that cannot be reduced to entailment. For example, we might like to say that if a sentence is a conjunction of two other sentences, then that is a logical property of that sentence. In that case, 'Snow is white' will have different logical properties from 'Snow is white and snow is white', even though these sentences entail each other.

[22] I mean it when I say that this is a start. Here is a better attempt. Subsentential expressions α and β have the same sense iff for any $\Phi(\gamma)$: (a) there exists a result of substituting α for γ in $\Phi(\gamma)$ iff there exists a result of substituting β for γ in $\Phi(\gamma)$; and (b) every result of substituting α for γ in $\Phi(\gamma)$ has the same sense as every result of substituting β for γ in $\Phi(\gamma)$. However, to keep the discussion simple, I will work with the rougher explication given in the main text; nothing will turn on this decision to keep things simple.

But if it really is possible for me to say everything I want to say without appearing to reify senses, then why don't I? Why do I insist on speaking as if we could refer to senses with singular terms? There are two answers to this question, one shallower and one deeper. The shallower answer is this: if I steadfastly refused to speak as if we could refer to senses with terms, this paper would become so unwieldy and complex that it would be impossible to understand. Even including all of the necessary explications in parentheses or footnotes would prove impractical, and so for the most part I will leave it as an exercise for the reader to come up with them as needed. The deeper answer is this: part of the point of this paper is to explicate some of our talk that apparently reifies senses. On the face of it, we can only talk about substituting senses if propositions are complex objects made up of senses; however, we can use my definition of sense-substitution to explicate this sort of talk without treating senses as objects. Crucially, though, I could not offer this explication unless I *started* by talking in the ordinary, unexplicated way about substituting senses. It is only by first displaying this unexplicated way of speaking that we can consider ways of explicating it.

4.2 The particular worry: my appeal to sense

In the last subsection I offered an explanation of what it is for two subsentential expressions to have the same sense, and that explanation was framed in terms of substitution: subsentential expressions α and β have the same sense if and only if substituting β for α in a sentence always produces a sentence with the same sense as the original, and *vice versa*.[23] But how should substitution be understood in this explanation? Not as simple-substitution. 'I' and 'me' may well have the same sense, but when we simple-substitute 'I' for 'me' in a sentence we usually end up with something ungrammatical, and hence something that does not have any sense at all. (Things only get worse when we consider synonymous expressions from different languages.) So we must be using a more sophisticated conception of substitution. And it seems as though only one conception will do: sense-substitution! But now, doesn't my definition of sense-substitution run in a circle? I defined sense-substitution in terms of subsentential expressions having the same sense, and then I explained what it is for subsentential expression to have the same sense in terms of sense-substitution.[24]

It might be possible to avoid this circularity, but the prospects on this front seem dim. There are only two strategies to try. First, we could try to offer a new explanation of what it is for two subsentential expressions to share a sense that does not proceed in terms of substitution. However, I simply have no idea of how to develop such an explanation, at least so long as we refuse to reify senses. Second, we could grant that any explanation of what it is for subsentential expressions to have the same sense will involve substitution, but deny that this needs to be understood as sense-substitution. But if not sense-substitution, then what? What

[23] My improved explanation in footnote 22 was also framed in terms of substitution, and so all of the following discussion applies to that explication as well.

[24] Thanks to Peter Sullivan for pushing this objection.

conception of substitution will fit the bill? Again, I do not know of any other kind of substitution which could do this job.[25]

It may be, then, that I have no choice but to live with the circularity in my definition of sense-substitution. Fortunately, however, that is something I think I can do. There are two points that I want to make in this connection. First, what this circularity shows us is that we can neither explain sense-substitution to someone who has no inkling of what it is for subsentential expressions to share a sense, nor explain what it is for subsentential expressions to share a sense to someone who has no inkling of sense-substitution. The proper response to this predicament is not to withdraw the definition of sense-substitution, but to reconsider the spirit in which we lay it down. We cannot hope to explain the notions of sense and sense-substitution to someone who has come to them cold; instead, all we can do is map the pre-existing relations that hold between these two notions. Taken in this spirit, my definition of sense-substitution is not an attempt to invent an all new notion of substitution, but to articulate a notion of substitution that was already, if only tacitly, essential to our notion of sense.

Second, the circularity in my definition of sense-substitution is most stark when we are speaking in a very general way about what it takes for two arbitrary expressions to have the same sense. But when we are discussing two particular expressions, we can sometimes mitigate this circularity by bringing to bear special assumptions about the senses of the expressions we are dealing with. This is a point that I will exploit throughout Sect. 6.[26]

4.3 A final remark about sense

I have now said everything of substance I have to say about the notion of sense used in this paper. But before ending this section, I would like to pause on a much more trivial point. I have so far spoken naïvely as if we can say once and for all whether one expression has the same sense as another. But of course, this is a simplification. An expression has a sense only relative to a given language, as used by a particular speaker, in a particular context, etc. etc. Nonetheless, it seems to me that simplicity is a virtue here. I will, then, continue to suppress these complexities, and I will do so by assuming that for the purposes of any given discussion, the senses of all the expressions under consideration are held fixed. (For the most part I will leave it to the reader to decide how the senses should be fixed.) I will also assume that when we hold their senses fixed, *has the same sense as* is an equivalence relation over meaningful expressions.

[25] I discuss one proposal in the "Appendix", and explain why I think it will not work.

[26] In an earlier paper (Trueman 2012a), I criticised Dolby's (2009) attempt to develop an account of substitution via quantification. I objected that Dolby's account was circular: in order to understand the quantificational moves that Dolby wanted to make, we already need the very notion of substitution that Dolby was trying to develop. At the time, I took it for granted that this circularity was vicious. Obviously I can no longer maintain that stance, not now that I have been forced to acknowledge a similar circularity in my own account.Nonetheless, I still think that Dolby's account compares unfavourably to my own. As we will see, we can put sense-substitution to substantial philosophical work, and it is not clear to me that the same is true of Dolby's notion of substitution. Unfortunately, however, I cannot pursue this matter any further here.

5 Everywhere and somewhere

Now that I have offered some defence of the sense in sense-substitution, I want to work towards an argument for a version of (RP). I will start by trying to argue for (RP_2),

(RP_2) Co-referring expressions are everywhere sense-intersubstitutable,

but as we will see, this is not quite the version of (RP) that we will end up with. I should also mention now that what I will present will really only be a *sketch* of an argument. At various points in this sketch, I will be forced to make more or less controversial assumptions, and although I will try to offer some justification for these assumptions, I simply will not have the space to make the case as watertight as I would like. What follows will not, therefore, establish once and for all that a version of (RP) is true; in the end, all I will do is present a way in which one might *try* to argue for a version of (RP).

In an ideal world, we would argue for (RP_2) simply by showing that it follows from the definition of sense-substitution, but unfortunately, that is something we cannot do. The definition of sense-substitution only tells us what is required for a sentence to count as a result of sense-substituting β for a given occurrence of α; it does not by itself tell us whether β *can* be sense-substituted for that occurrence of α. We will, then, need something more than the definition of sense-substitution to get an argument for (RP_2). To that end, I would like to introduce the following principle:

($*$) If γ can be sense-substituted for α in $\Phi(\alpha)$ but not for β in $\Psi(\beta)$, then $\alpha \not\approx \beta$.

We can offer the following argument for ($*$):

> Suppose that γ can be sense-substituted for α in $\Phi(\alpha)$, but not for β in $\Psi(\beta)$. Sense-substitution is a way of substituting one sense for another. The sense of a subsentential expression is its contribution to the senses of the sentences in which it appears. So, sense-substitution is a way of substituting these contributions. Thus, it is possible to substitute the contribution made by γ for the contribution made by α to the sense of $\Phi(\alpha)$, but impossible to substitute the contribution made by γ for the contribution made by β to the sense of $\Psi(\beta)$. By Leibniz's Law it follows that the contribution made by α to the sense of $\Phi(\alpha)$ is distinct from the contribution made by β to the sense of $\Psi(\beta)$. In other words, the sense of α in $\Phi(\alpha)$ is distinct from the sense of β in $\Psi(\beta)$. And since we are holding the senses of expressions fixed for the purposes of any given discussion, we can say more simply: the sense of α is not the sense of β.[27]

[27] An anonymous reviewer (the same one as before) pointed out to me that if this argument works, then it can easily be re-worked to show the following:

> If we can produce a *true* sentence by sense-substituting γ for α in $\Phi(\alpha)$, and a false sentence by sense-substituting γ for β in $\Phi(\beta)$, then $\alpha \not\approx \beta$.

However, this further principle is by no means obviously correct: on the face of it, '*a* is a vixen because *a* is a female fox' is true and '*a* is a vixen because *a* is a vixen' is false, even though (indeed, precisely

On the face of it, accepting this argument requires thinking of senses as objects: all the way through it we spoke as if senses could be referred to with singular terms—we even invoked Leibniz's Law! But, as I have emphasised several times now, I do not want to assume that senses are objects in this paper. Nonetheless, I think that we can still find this argument compelling, so long as we take it in the right spirit. What it unproblematically shows is that according to our pre-explication thought about senses, in which we think of senses as objects, the following claim is true:

(†) If the sense of γ can be substituted for the sense of α in the proposition expressed by $\Phi(\alpha)$ but not for the sense of β in the proposition expressed by $\Phi(\beta)$, then the sense of α is not the sense of β.

As I mentioned in Sect. 4.1, sense-substitution is meant to give us a way of explicating our talk of substituting one sense for another *without reifying senses*; this is how we should understand the claim that sense-substitution is a way of substituting one sense for another. So, $(*)$ should itself be seen as a way of explicating (†) without trying to reify senses. In summary, then, the above argument works by establishing that (†) is part of our pre-explication thought about senses, and that $(*)$ is a way of explicating (†).

Let us from now on take $(*)$ as given. An immediate corollary of $(*)$ is that if γ can be sense-substituted for *some* occurrence of α, then γ can be sense-substituted for *every* occurrence of α. To see this, just let β in $(*)$ be identical to α:

If γ can be sense-substituted for α in $\Phi(\alpha)$ but not for α in $\Psi(\alpha)$, then $\alpha \not\approx \alpha$.

Since we are assuming that *has the same sense as* is an equivalence relation over meaningful expressions, α must have the same sense as α. Hence we can infer that if γ can be sense-substituted for α in $\Phi(\alpha)$, then it can also be sense-substituted for α in $\Psi(\alpha)$.

Straightaway, then, $(*)$ makes our task easier. All we have to do to show that a pair of co-referring expressions are *everywhere* sense-intersubstitutable is show that they are *somewhere* intersubstitutable. Let us focus again on 'Bertrand' and 'the referent of "Bertrand" '. Are these two terms somewhere sense-intersubstitutable? Well, consider the following sentences:

(4a) 'Bertrand' refers to Bertrand
(4b) 'Bertrand' refers to the referent of 'Bertrand'.

(4a) and (4b) are both meaningful; in fact they are both true. Moreover, they involve no illicit pun: ' "Bertrand" refers to ( )' has the very same sense in (4a) and (4b).

Footnote 27 continued
because) '( ) is a vixen' has the same sense as '( ) is a female fox'. Unfortunately, I do not have the means to address this kind of hypersensitivity in this paper (and that is not just because I lack the space!). For now, all I can do is mention that I am sceptical of the idea that when 'because' is used in this hypersensitive way, sentences of the form '*p* because *q*' have any truth-conditions at all; rather than saying something true, I suspect that '*a* is a vixen because *a* is a female fox' acts as a kind of conceptual aid, an instruction to connect up our concepts in a certain way.

So, 'Bertrand' and 'the referent of "Bertrand" ' are *somewhere* sense-intersubstitutable, and thus by $(*)$, they are *everywhere* sense-intersubstitutable.

At this point we might wonder whether this argument could be generalised into an argument for (RP_2). (4a) is an instance of the disquotation schema,

(D) '*a*' refers to *a*,

and it is generally agreed that disquotation is somehow essential to reference. So couldn't we argue as follows?

> Suppose that '*c*' and '*d*' co-refer. Since disquotation is essential to reference, and '*c*' and '*d*' are referring expressions, both of the following instances of (D) must be true:
>
> (5a) '*c*' refers to *c*.
> (5b) '*d*' refers to *d*.
>
> As '*c*' and '*d*' co-refer, we can infer from (5a) and (5b) that the following is true:
>
> (5c) '*c*' refers to *d*.
>
> Now, if (5a) and (5c) are both true, then they must also both be meaningful. So (5c) is a result of sense-substituting '*d*' for '*c*' in (5a). Thus '*c*' and '*d*' are somewhere sense-intersubstitutable, and hence by $(*)$, they are everywhere sense-intersubstitutable.

Unfortunately, however, this argument will not stand up to scrutiny. As we will see, there is a problem with the way in which it appeals to disquotation, and although there is a way around that problem, it leads us to another new version of (RP). But before we can get to any of that, we need to pause on a preliminary discussion about the sense-substitution of predicates.

6 Sense-substitution and predicates

We need some more terminology and notation. A *monadic first-level* predicate is what you get when you remove a singular term from a sentence. More generally, an *n*-adic first-level predicate is what you get when you remove *n* singular terms from a sentence, but to keep things simple, I will focus on the monadic case. I will, then, mean monadic predicates by 'predicate', unless I clearly indicate otherwise. We can also remove first-level predicates from a sentence; when we do, we thereby produce *second-level* predicates. Again, I will keep things simple by focussing on first-level predicates, and so I will mean first-level predicates by 'predicate', unless I clearly indicate otherwise. Since it can be useful to visually differentiate first-level predicates from second-level ones, I will mark the gaps that result from removing singular terms with a sans-serif 'x' or 'y', and the gaps that result from removing monadic first-level predicates with a sans-serif 'X' or 'Y'. So, for example, when we remove the term 'Shergar' from 'Shergar is a horse', we get the first-level predicate

‘x is a horse’, and when we remove ‘x is a horse’ from ‘$\exists x$(*x* is a horse)’, we get the second-level predicate ‘$\exists x(\mathsf{X}x)$’.[28]

So far in this paper, we have only considered examples of sense-substituting one singular term for another. But we can also sense-substitute one predicate for another. Take for example,

(6a) London is a city,

and imagine that we wanted to sense-substitute ‘x is crowded’ for ‘x is a city’. One obvious choice is:

(6b) London is crowded.

(6b) counts as a result of sense-substituting ‘x is crowded’ for ‘x is a city’ in (6a) if,

(a) ‘X(London)’ ≈ ‘X(London)’
and (b) ‘x is crowded’ ≈ ‘x is crowded’,

both of which are trivial, given our assumption that ≈ is an equivalence relation. But again, (6b) is not the only result. The French sentence,

(6c) Londres est bondé,

also counts as a result of this sense-substitution so long as,

(a) ‘X(Londres)’ ≈ ‘X(London)’
and (b) ‘x est bondé’ ≈ ‘x is crowded’,

both of which look plausible.

We can, then, sense-substitute one predicate for another. But what we cannot do is sense-substitute a predicate for a singular term. Suppose, for example, that we wanted to sense-substitute ‘x is crowded’ for ‘London’ in (6a). To do so, we would need to construct a meaningful sentence $\Psi(\gamma)$ such that:

(a) $\Psi(\ ) \approx$ ‘x is a city’
and (b) $\gamma \approx$ ‘x is crowded’.

(a) and (b) tell us that $\Psi(\ )$ and γ both have the senses of first-level predicates. But we cannot construct a meaningful sentence by writing an expression with the sense of a first-level predicate into the gap in an expression that *also* has the sense of a first-level predicate. First-level predicates are what you get when you remove singular terms from sentences, and their senses reflect that fact. Roughly, singular terms pick out objects, and first-level predicates say things of objects; for example, in ‘London is a city’, ‘London’ picks out an object — London—and ‘x is a city’ says something of that object—that it is a city. But obviously, we cannot make the senses of two first-level predicates work together in anything like this way. That is

[28] To be absolutely clear: these sans-serif letters are not variables or schematic letters; they merely mark the gaps in predicates. Frege used Greek letters like ‘ξ’ and ‘ζ’ rather than these sans-serif letters, but I am already using Greek letters as metalinguistic variables ranging over expressions. And anyway, using the same letters in different fonts a gap markers and as variables has the added benefit of allowing us to move smoothly between them.

why I have not marked the type of gap in $\Psi(\ )$: (a) tells us that $\Psi(\ )$ is a first-level predicate, $\Psi(\mathsf{x})$, but (b) tells us that it cannot be.

It is worth pre-empting two possible misunderstandings here. First, we might object that in the foregoing discussion, I must have assumed that predicates are not referring expressions. Otherwise, I would have seen no obstacle to constructing a meaningful sentence by writing a first-level predicate into the gap of another first-level predicate: in the sentence $\Psi(\gamma)$, γ picks out a particular property—the property of being crowded—and $\Psi(\ )$ says something of that property—that it is a city.[29] But I made no such assumption. The problem with $\Psi(\gamma)$ was that since γ has the sense of a predicate, at least part of its job is to say something of an object, and that is something it cannot do in $\Psi(\gamma)$. This point still holds even if predicates are also referring expressions.[30] Indeed, the proposed way of understanding $\Psi(\gamma)$ illustrates that fact: clearly, γ does not there serve to say something of anything.

Second, the above remarks are not meant to suggest that the *only* way to construct a meaningful sentence with a first-level predicate is to fill its gap with a singular term. Obviously, we can also write a first-level predicate into the gap in a second-level predicate. But crucially, this is possible only because second-level predicates have a different kind of sense from first-level predicates. Second-level predicates are what you get when you remove first-level predicates from sentences, and their senses reflect that fact. One of the things that second-level predicates do is supply domains of objects for first-level predicates to say things of. Quantifiers are the paradigms here. The first-order quantifiers '$\exists x(\mathsf{X}x)$' and '$\forall x(\mathsf{X}x)$' are second-level predicates, and we are all familiar with the idea that these quantifiers come with a domain of quantification: '$\exists x(x$ is a horse)' is true just in case 'x is a horse' says something true of *some* object in the domain, and '$\forall x(x$ is a horse)' is true just in case 'x is a horse' says something true of *every* object in the domain. But all second-level predicates can be said to supply domains of objects for first-level predicates to say things of. We can even say that 'X(London)', which is a sort of limiting case of a second-level predicate, supplies such a domain, albeit one that includes only London: when we plug 'x is a city' into 'X(London)' we get 'London is a city', which is true just in case 'x is a city' says something true of the object in the domain supplied by 'X(London)'.[31]

In the last few paragraphs I have made a number of claims about how terms and predicates function, about the different kinds of sense they enjoy.[32] I hope that these

[29] It seems to me that this line of thought is present in Magidor's (2009: pp. 6–7).

[30] For the record, I do think predicates refer, but I do not think that they refer in the same sense as singular terms; in fact, I think that *what it is* for a predicate to refer is for it to say something of an object. See Trueman (2015) and Sect. 7–9 of this paper. For an alternative conception of predicate reference, see MacBride's (2011: esp. Sect. 4).

[31] And in case there is any doubt, the same can also be said in more complex cases. For example, the second-level predicate '$Nx(\mathsf{X}x)$ is even' supplies a domain of objects for first-level predicates to say things of: '$Nx(x$ is a horse) is even' is true just in case 'x is a horse' says something true of an even number of objects in that domain.

[32] In particular, I have been working with a broadly Fregean conception of terms and predicates. See Sullivan's (2010) for an excellent explanation of what is involved in that conception, although Sullivan (pp. 116–117) ultimately expresses some reservations about it.

claims are relatively uncontroversial, but I am sure that they are not so uncontroversial that no philosopher would controvert them.[33] Nonetheless, in what follows I will take it for granted that the above reasoning is accepted, and thus that predicates cannot be sense-substituted for singular terms.

7 Disquotation and reference

At the end of Sect. 5, I outlined an argument that we might try to give for (RP_2). That argument relied on the assumption that we always get a true sentence when we substitute a referring expression into the following schema:

(D) '*a*' refers to *a*.

That assumption was meant to be justified by the thought that disquotation is somehow essential to reference. Now, I do not want to deny that there is some kind of essential relationship between disquotation and reference. (Indeed, the main aim of this section is to get clearer on what that relationship consists in.) Nonetheless, it is important to note that substituting a referring expression into (D) does not always yield a true sentence. Consider the following failed instances of (D):

(7a) 'I' refers to I
(8a) 'Londres' refers to Londres
(9a) '**x** is a horse' refers to is a horse.

'I' and 'Londres' are certainly referring expressions, and many philosophers would say that '**x** is a horse' is as well, yet none of (7a)–(9a) are grammatical, let alone true.

What should we say about these failed instances of (D)? Well, I hope that by now it is obvious how to deal with (7a). Clearly, the problem here is that we *simple*-substituted 'I' for the second '*a*' in (D), and all we need to do to get around this problem is *sense*-substitute 'I' for that '*a*' instead. In other words, we just need to sense-substitute 'I' into the gap in ' "I" refers to ( )'.[34] If we continue to assume that 'I' and 'me' have the same sense, then

(7b) 'I' refers to me,

will count as a result of this sense-substitution. What is more, (7b) is true, or at least it is when the context is held fixed in the appropriate sort of way.[35]

[33] Ramsey's (1925: esp. 116) is sometimes read as a rejection of my broadly Fregean assumptions about terms and predicates; however, this interpretation has been challenged by MacBride (2005).

[34] I am speaking slightly loosely here. As I emphasised in Sect. 3, sense-substitution is defined only in relation to meaningful whole sentences. This loose way of speaking is, however, harmless: to say that α can be sense-substituted into the gap in $\Phi(\ )$ is to say that α can be sense-substituted for β in any meaningful sentence $\Psi(\beta)$ where $\Psi(\ ) \approx \Phi(\ )$.

[35] That is, (7b) will be true if I utter it and am saying what 'I' refers to when I use it, or if you utter it and are saying what 'I' refers to when you use it, and so on.

We can deal with (8a) in much the same way. (8a) is the result of simple-substituting the French term 'Londres' into the English schema (D), and so it is hardly surprising that it is ungrammatical. However, recalling a point made in Sect. 3, the fact that 'Londres' is French and (D) is English does not mean that we cannot produce a true sentence by *sense*-substituting 'Londres' into the gap in ' "Londres" refers to ( )'. Both of these sentences will surely do:

(8b) 'Londres' refers to London
(8c) 'Londres' se réfère à Londres.

But now we come to (9a), which is a much more difficult sort of case. Here we cannot deal with the problem simply by moving over to sense-substitution. Ordinarily understood, 'refers to' is a first-level predicate, 'x refers to y'. For example, in the sentence ' "Bertrand" refers to Bertrand', 'refers to' is flanked by ' "Bertrand" ', a term referring to the term 'Bertrand', and 'Bertrand', a term referring to the man Bertrand. So the gap left open in ' "x is a horse" refers to ( )' is a gap for singular terms: ' "x is a horse" refers to y'. But as we saw in Sect. 6, predicates cannot be sense-substituted for singular terms, and so 'x is a horse' cannot be sense-substituted into that gap.

It might be tempting to say that all this shows us is that 'x is a horse' does not refer to anything: disquotation is essential to reference, and so if we cannot disquote 'x is a horse' then it cannot be a referring expression. And had the problem been that we end up with a false sentence when we sense-substitute 'x is a horse' into the gap in ' "x is a horse" refers to y', then this surely would have been the right response. However, the problem is much more radical than that: no sentence, true or false, results from sense-substituting 'x is a horse' into that gap. And as a result, I think, a much more radical response is called for. We must draw a sharp distinction between two kinds of reference: one for singular terms, expressed by 'x refers to y', and one for predicates, expressed by something of the form '$R(x, Y)$'. I have argued for this more radical response elsewhere,[36] but I cannot reproduce that argument in all of its detail here. Nonetheless, it would still be useful for me to try to give some sense as to why I think that this more radical response is called for.

Let's temporarily set 'x is a horse' to one side, and consider the empty singular term 'Vulcan'. Now, since 'Vulcan' is a term, we can sense-substitute it into the gap in ' "Vulcan" refers to y' without any difficulty: ' "Vulcan" refers to Vulcan' will do. Of course, since 'Vulcan' is an empty term, ' "Vulcan" refers to Vulcan' is false.[37] Nonetheless, it is still true to say: If 'Vulcan' refers to something, then 'Vulcan' refers to Vulcan. Indeed, that seems to be almost analytically true.[38] To

[36] See Trueman (2015), Sect. 5.

[37] Some philosophers (e.g. Frege 1892: pp. 156–157, 1897: pp. 229–230) have insisted that sentences featuring empty names are not truth-evaluable. Unfortunately, this is yet another issue that lies beyond the scope of this paper.

[38] I say '*almost* analytically true' because analytic truths are normally thought to be necessarily true, and there are certainly readings of 'If "Vulcan" refers to something, then "Vulcan" refers to Vulcan' on which it is not necessarily true: we could have used the word 'Vulcan' as a name for Mars.

see this, suppose that ‘ “Vulcan” refers to something’ is true, but ‘ “Vulcan” refers to Vulcan’ is not. In that case, the predicate ‘ “Vulcan” refers to y’ would have to be satisfied by something *other than* what we refer to when we use ‘Vulcan’. Our use of ‘ “Vulcan” refers to y’ would thus be cut free from our use of ‘Vulcan’ itself, and as a result, it would no longer be possible to use ‘ “Vulcan” refers to y’ to specify what ‘Vulcan’ refers to.

Now let’s return to ‘x is a horse’. As I said earlier, the problem is not merely that we cannot produce a *true* sentence by sense-substituting ‘x is a horse’ into the gap in ‘ “x is a horse” refers to y’. The problem is that we cannot sense-substitute ‘x is a horse’ into that gap at all. Consequently, we cannot sense-substitute ‘x is a horse’ into the gap in ‘If “x is a horse” refers to something, then “x is a horse” refers to y’ either. Now, we just said that if we could not produce a true sentence by sense-substituting ‘Vulcan’ into the gap in ‘If “Vulcan” refers to something, then “Vulcan” refers to y’, then we could not use ‘ “Vulcan” refers to y’ to specify what ‘Vulcan’ refers to. By analogy, then, we must surely conclude that we cannot use ‘ “x is a horse” refers to y’ to specify what ‘x is a horse’ refers to: our use of ‘ “x is a horse” refers to y’ simply is not tethered in the right way to our use of ‘x is a horse’ itself.

This may at first seem like a bizarre thing to say, but what it really comes down to is just this: we cannot talk about predicate reference in the same way we talk about term reference; instead, we need to use a notion of reference that is tailor made for predicates. In particular, we need to use a second-level version of reference, expressed by a predicate of the form ‘$R(\mathsf{x}, \mathsf{Y})$’; the first gap here, marked ‘x’, is a gap for a term referring to a predicate,[39] and the second gap, marked ‘Y’, is a gap into which a predicate can be sense-substituted.[40] Now, there is a good question about which predicate of the form ‘$R(\mathsf{x}, \mathsf{Y})$’ we should use to express this second-level version of reference, but for present purposes it does not matter which one we choose.[41] For now, then, I will just use ‘x predicate-refers to Y’ as a new primitive which is to express the notion of reference for predicates. (And when it is helpful, I will also write ‘x term-refers to y’ instead of ‘x refers to y’.) We can now disquote ‘x is a horse’ as follows:

(9b) ‘x is a horse’ predicate-refers to is a horse.

(9b) might look ungrammatical as a piece of English, but ‘x predicate-refers to Y’ is a new addition to English, and (9b) is grammatical by design.

[39] I am here assuming that predicates are objects and so capable of being referred to by singular terms. This assumption may well be mistaken, but nothing in my argument essentially turns on it: I want to focus on the second gap in ‘$R(\mathsf{x}, \mathsf{Y})$’, not the first.

[40] I have just stretched my definition of sense-substitution even further. What does it mean to say that α can be sense-substituted into the second of the *two* open gaps in $\Phi(\ )_1(\ )_2$? This will do for now: α can be sense-substituted for γ in any meaningful sentence $\Psi(\beta)_1(\gamma)_2$ where $\Psi(\ )_1(\ )_2 \approx \Phi(\ )_1(\ )_2$. This definition will no doubt need some work to make it generally applicable, but it will suffice for the cases discussed below.

[41] For the record, I would like to suggest that ‘$\forall y(y \text{ satisfies } \mathsf{x} \leftrightarrow \mathsf{Y}y)$’ is the right choice here, or at least I would if we could agree to set intensional contexts aside for the time being. However, nothing in this paper turns on this suggestion.

As I warned, this is nothing more than the outline of an argument; it takes a great deal of work to fill in all of the details. But in what remains, I will assume that its conclusion is accepted: the claim that '**x** is a horse' refers should be understood as the claim that '**x** is a horse' predicate-refers. However, it is again worth pre-empting a possible misunderstanding here. Nowhere in this section have I argued that '**x** is a horse' actually predicate-refers to anything. In fact, for all I have said here, it may be that no predicate predicate-refers to anything.[42] The point was this. We need to understand the claim that '**x** is a horse' refers as the claim that '**x** is a horse' predicate-refers so that we can meaningfully disquote '**x** is a horse' with (9b). But the fact that (9b) is meaningful does not by itself tell us that '**x** is a horse' predicate-refers to anything, any more than the fact that ' "Vulcan" term-refers to Vulcan' is meaningful tells us that 'Vulcan' term-refers to anything. '**x** is a horse' predicate-refers to something only if (9b) is not just meaningful but *true* as well.

8 One last Reference Principle

In Sect. 6 I argued that predicates cannot be sense-substituted for singular terms. (RP_2) would, therefore, have us conclude that terms and predicates do not co-refer. But then, in Sect. 7, I argued that we need to distinguish between two notions of reference: one for singular terms, expressed by '**x** term-refers to **y**', and one for predicates, expressed by '**x** predicate-refers to **Y**'. And when we do, we reach a far more striking conclusion than the one issued by (RP_2): rather than being false, it is just nonsensical to say that a singular term term-refers to the same thing that a predicate predicate-refers to.

If we wanted to say that a singular term, '*c*', term-refers to the same thing that a predicate, '*F***x**', predicate-refers to, we would need to bind both the variables in,

(10) '*c*' term-refers to *y* and '*F***x**' predicate-refers to *Y*,[43]

with a single existential quantifier. But that is not something we can do. Just imagine trying to form an instance of this generalisation. We would have to substitute the one expression for both the variables in (10). Now, the notion of substitution people usually have in mind when they talk about forming an instance of a generalisation is simple-substitution, but in the current context that seems wrongheaded. Better to use sense-substitution instead. So, we would need to sense-substitute one expression for both the variables in (10). But no *one* expression can be sense-substituted for *both* of these variables: we can sense-substitute a term but not a predicate into the gap in ' "*c*" term-refers to **y**', and we can sense-substitute a predicate but not a term into the gap in ' "*F***x**" predicate-refers to **Y**'. It is not, therefore, possible to form an instance of the generalisation that would result from

[42] Nonetheless, if my earlier suggestion that we should read '**x** predicate-refers to **Y**' as '$\forall y$(*y* satisfies **x** $\leftrightarrow$ **Y***y*)' is right, then we cannot coherently deny that predicates refer. But that is a discussion for another time.

[43] For added concreteness, you could imagine that '*F***x**' is '**x** is a horse' and '*c*' is 'the property of being a horse'.

binding both the variables in (10) with a single existential quantifier,[44] and that surely just means that it is not possible to form that generalisation in the first place.

Once again, it is worth pre-empting a possible misunderstanding here. It is tempting to say that an existential generalisation is *true* only if we can produce a *true* instance of that generalisation. But whether or not that thought is right,[45] it was not the thought underwriting the argument above. Rather, my thought was that we cannot form a *meaningful* generalisation, true or false, unless we can produce a *meaningful* instance of that generalisation. And that is a thought that I cannot imagine gainsaying.

So it is nonsense, not false, to say that a term term-refers to the same thing that a predicate predicate-refers to. This suggests that rather than working with (RP_2), we should move over to another new version of (RP). To formulate this new principle, we will need some new notation. From now on, I will use '**x** refers$_\alpha$ to ( )' to express the appropriate notion of reference for α; the first gap here, marked '**x**', is a gap for a term referring to α, and the second gap, whose type I have not marked, is a gap into which α can be sense-substituted. The idea behind this notation is a generalisation of what I said about predicate-reference in Sect. 7. I argued that if we want to say the predicate '**x** is a horse' does or does not refer to something, we need to use a notion of reference that allows us to disquote '**x** is a horse'. But there is nothing special about predicates: in general, then, if we want to say that an expression α does or does not refer to something, we must use a notion of reference that allows us to disquote α.[46]

We can now present our new version of (RP):

(RP_3) If it makes sense to say that α refers$_\alpha$ to what β refers$_\beta$ to, then α and β are everywhere sense-intersubstitutable.[47]

What is more, we can actually offer an argument for (RP_3). This argument has three (schematic) premises:[48]

[44] Some nominalists, who do not think that predicates refer, would say that it is impossible to bind the variable '*Y*' in ' "*F***x**" predicate-refers to *Y*' with a quantifier at all. However, in Trueman (2012b), I argue that even if predicates do not refer, we can still quantify into predicate position.

[45] It turns out that this thought is a little more problematic than it might seem. See Levine (2013) for an interesting discussion.

[46] In case there is any confusion, I am not suggesting that every expression has its own unique notion of reference. If α and β are everywhere sense-intersubstitutable, then I see no reason to deny that '**x** refers$_\alpha$ to ( )' and '**x** refers$_\beta$ to ( )' both express the same notion of reference.

[47] Unlike (RP_1) and (RP_2), (RP_3) must be read as a schema rather than a universal generalisation. That is because 'α' and 'β' both play two roles in (RP_3): they appear in the positions of terms and in undetachable subscripts. In particular, (RP_3) is intended to be a schematic generalisation over *meaningful* expressions, i.e. expressions which have senses.

[48] Like (RP_3), Premises 2 and 3 must be read as schematic generalisations. And for the purposes of the following argument, (∗) may as well be a schema too.

Premise 1

($*$) If γ can be sense-substituted for α in $\Phi(\alpha)$ but not for β in $\Psi(\beta)$, then $\alpha \not\approx \beta$.

Premise 2

α can be sense-substituted into the second gap in 'x refers$_\alpha$ to ( )'.

Premise 3

If it makes sense to say that α refers$_\alpha$ to what β refers$_\beta$ to, then it is possible to sense-substitute a single expression into the gaps in both $\ulcorner\mathcal{A}$ refers$_\alpha$ to ( ) $\urcorner$ and $\ulcorner\mathcal{B}$ refers$_\beta$ to ( ) $\urcorner$, where $\mathcal{A}$ is some term referring to α and $\mathcal{B}$ is some term referring to β.[49]

Premise 1 is, of course, just ($*$) from Sect. 5. Premise 2 was built into the notation of 'x refers$_\alpha$ to ( )': 'x refers$_\alpha$ to ( )' can express the appropriate notion of reference for α only if we can sense-substitute α into its second gap.[50] Premise 3 is an application of the idea that we can form a generalisation only if it is possible to form an instance of that generalisation. With these three premises, we can argue as follows:

Suppose that it makes sense to say that α refers$_\alpha$ to what β refers$_\beta$ to. It follows from Premise 3 that we can sense-substitute one expression γ into the gaps in both $\ulcorner\mathcal{A}$ refers$_\alpha$ to ( ) $\urcorner$ and $\ulcorner\mathcal{B}$ refers$_\beta$ to ( ) $\urcorner$. Therefore, there is some possible sentence $\Phi(\phi)$ such that,

(a) $\Phi(\ ) \approx \ulcorner\mathcal{A}$ refers$_\alpha$ to ( ) $\urcorner$
and (b) $\phi \approx \gamma$,

and there is another possible sentence $\Psi(\psi)$ such that,

(a) $\Psi(\ ) \approx \ulcorner\mathcal{B}$ refers$_\beta$ to ( ) $\urcorner$
and (b) $\psi \approx \gamma$.

By Premise 2 it is possible to sense-substitute α into the gap in $\ulcorner\mathcal{A}$ refers$_\alpha$ to ( ) $\urcorner$, and thus there is a possible sentence $\Sigma(\sigma)$ such that:

(a) $\Sigma(\ ) \approx \ulcorner\mathcal{A}$ refers$_\alpha$ to ( ) $\urcorner \approx \Phi(\ )$
and (b) $\sigma \approx \alpha$.

[49] The corner-quotes in this premise are Quine-quotes, which we cannot omit this time. Since Quine-quotes can be a headache, it might be useful to give an example. Let α be 'Bertrand'. $\mathcal{A}$ needs to be a term referring to α, so let $\mathcal{A}$ be ' "Bertrand" '. And let us abbreviate the reference-predicate for 'Bertrand' as 'x refers$_B$ to ( )'. With all this in place, $\ulcorner\mathcal{A}$ refers$_\alpha$ to ( ) $\urcorner$ is ' "Bertrand" refers$_B$ to ( )', and $\ulcorner\mathcal{A}$ refers$_\alpha$ to $\alpha\urcorner$ is ' "Bertrand" refers$_B$ to Bertrand'.

[50] Couldn't there be a meaningful expression, α, for which there is no appropriate notion of reference? Maybe. But in that case, the relevant instances of (RP_3) will be vacuous: 'x refers$_\alpha$ to ( )' will be meaningless, and so it will not make sense to say that α refers$_\alpha$ to what β refers$_\beta$ to. In what follows, then, I will make the more interesting assumption that there are appropriate notions of reference for the expressions under discussion.

Similarly, by Premise 2 it is possible to sense-substitute β into the gap in $\ulcorner\mathcal{B}$ refers$_\beta$ to $(\ )\urcorner$, and thus there is a possible sentence $\Delta(\delta)$ such that:

(a) $\Delta(\ ) \approx \ulcorner\mathcal{B}$ refers$_\beta$ to $(\ )\urcorner \approx \Psi(\ )$
and (b) $\delta \approx \beta$.

It is now trivial that we can sense-substitute α for ϕ in $\Phi(\phi)$: $\Sigma(\sigma)$ counts as a result of this sense-substitution. And since $\phi \approx \gamma \approx \psi$, it follows by $(*)$ that we can sense-substitute α for ψ in $\Psi(\psi)$. So there is some possible sentence $\Lambda(\lambda)$ such that:

(a) $\Lambda(\ ) \approx \Psi(\ ) \approx \ulcorner\mathcal{B}$ refers$_\beta$ to $(\ )\urcorner$
and (b) $\lambda \approx \alpha$

Since $\Lambda(\ ) \approx \ulcorner\mathcal{B}$ refers$_\beta$ to $(\ )\urcorner \approx \Delta(\ )$, $\Lambda(\lambda)$ also counts as a result of sense-substituting α for δ in $\Delta(\delta)$. And as $\delta \approx \beta$, it follows by $(*)$ that α can be sense-substituted for every occurrence of β.

So, from the supposition that it makes sense to say that α refers$_\alpha$ to what β refers$_\beta$ to, we can infer that α can be sense-substituted for every occurrence of β; and of course, exactly the same reasoning will also allow us to infer from this supposition that β can be sense-substituted for every occurrence of α. Thus we have (RP_3).

9 Conclusion

I want to end now by returning to the easy arguments with which we started. On the face of it, we seem to have gone a long way towards vindicating them: they used (RP) to convert seemingly difficult questions into trivial ones, and we have just argued for a version of (RP). However, this initial assessment is quite mistaken.

To begin with, part of what made those easy arguments seem so easy was that on first glance, (RP) appeared to be a truism. But by now I hope that that appearance has been completely dispelled. While we do have a sketch of an argument for (RP_3), that sketch is by no means trivial. I was forced to make a number of controversial assumptions during the course of arguing for (RP_3), and I cannot pretend that I have given those assumptions a full defence here. Probably most notable in this regard is my assumption that we must distinguish between different kinds of reference for different kinds of expression. Everyone agrees that disquotation is *somehow* essential to reference, but I am sure that plenty of philosophers will be resistant to the idea that if we want to say that α does or does not refer to something, then we must use a notion of reference that allows us to disquote α. Ultimately, I do not believe that this idea can be resisted, but it would be absurd for me to suggest that I have done enough to show that here. All I have done is present a way of trying to argue for (RP_3), and while that is not nothing, it is nowhere near the last word on the subject.

But let us set that issue to one side and just take (RP_3) for granted. Still, the supposedly 'easy' arguments are not all that easy: it can be very hard to figure out whether two expressions are everywhere sense-intersubstitutable. This is a real difference between (RP_1) and (RP_3). Simple-substitution is a mechanical procedure that yields unique results. If we want to argue that β cannot be simple-substituted *salva congruitate* for α in $\Phi(\alpha)$, then all we really need to do is write out $\Phi(\beta)$; after that, anyone fluent in the relevant language can see whether we are right. Sense-substitution, on the other hand, is not mechanical and does not yield unique results. If we want to argue that β cannot be sense-substituted for α in $\Phi(\alpha)$, then we need to somehow show that *no possible* sentence could count as a result of this sense-substitution.

It is, then, pretty difficult to show that two expressions are not everywhere sense-intersubstitutable. But that is not to say that it is impossible. In Sect. 6 I tried to give some reason for thinking that predicates cannot be sense-substituted for terms. I think, then, that we can recover something from Easy Argument 1. That argument tried to use (RP_1) to show that the predicate 'x is a horse' does not co-refer with the term 'the property of being a horse'. We can now use (RP_3) to show that it is not even false to say that 'x is a horse' and 'the property of being a horse' co-refer; it is just nonsense. (Although to my mind this gives us no reason to deny that predicates refer in their own way.) But to repeat, what we have here can no longer be called an 'easy' argument. It is not just given that terms and predicates are not everywhere sense-intersubstitutable; establishing that result requires substantial philosophical work, work that I have only made a start on in this paper.

It is a lot harder to recover anything from Easy Argument 2. That argument tried to use (RP_1) to show that 'that the show will be funny' does not co-refer with 'the proposition that the show will be funny'. If we wanted to replace (RP_1) in that argument with (RP_3), we would need to find some way of arguing that 'that the show will be funny' and 'the proposition that the show will be funny' are not everywhere sense-intersubstitutable. Now, as it happens, I think that it is possible to supply such an argument,[51] but there is no denying that this is still very much a matter for debate. Indeed, for the time being we should even be open to the possibility that (in at least some of its uses) 'that the show will be funny' has exactly the same sense as 'the proposition that the show will be funny'; it may be that the grammatical differences between these expressions do not reflect differences in their senses, just as it may be that the grammatical differences between 'I' and 'me' do not reflect differences in their senses. If that turned out to be right, then 'that the show will be funny' and 'the proposition that the show will be funny' would be trivially sense-intersubstitutable. It takes a lot of work, then, to decide whether (RP_3) can be brought to bear on this case; indeed, the work that needs to be done is of the very kind that Easy Argument 2 was meant to excuse us from.

[51] See Trueman (2017).

In the end it does not seem that we can use (RP) to subserve any easy arguments, and really, that is what we should have expected. (RP_1) operates on the surface of language, which is why it is so easy to use. It invites us to draw substantial philosophical conclusions from the surface features of particular natural languages. But natural languages have too many quirks and idiosyncrasies at surface level to be allowed the final say on philosophical matters. (RP_3), on the other hand, runs much deeper. It will not allow us to draw any philosophical conclusions until we have shown that certain claims cannot be expressed in *any* language whatsoever, and that is not an easy thing to do. Here I can do no better than to borrow a line from Ramsey:

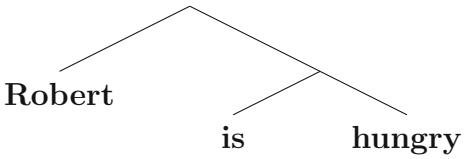
> Let us remind ourselves that the task on which we are engaged is not merely one of English grammar; we are not school children analysing sentences into subject, extension of the subject, complement and so on, but are interested not so much in sentences themselves, as in what they mean, from which we hope to discover the logical nature of reality. (Ramsey 1925: p. 117)

Acknowledgments I would like to thank Tim Button, Nick Jones, Michael Price and Peter Sullivan for all of their help with this paper. Thanks also to the anonymous reveiwer for this journal, who supplied a very helpful set of comments.

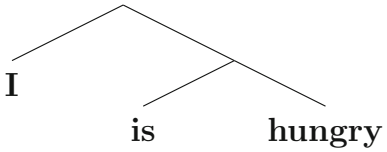
Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

In an exceptionally helpful set of comments, an anonymous reviewer for this journal presented an alternative conception of substitution. Suppose we wanted to substitute 'I' for 'Robert' in 'Robert is hungry'. On the anonymous reviewer's account, the first thing we do is step down from the surface-level sentence 'I am hungry', to the corresponding *phrase structure tree*, which looks something like this:



(The expressions written in bold here are lexical items, where ***a*** is the lexical item corresponding to '*a*'.) We then simple-substitute **I** for **Robert** in this tree:



We then step back up from this tree to the corresponding surface-level sentence. Importantly, this step back up will often involve some syntactic transformations. In this particular case, we must transform the verb to make sure that it agrees with the grammatical person of **I**. So what we end up with on the surface-level is not 'I is hungry', but 'I am hungry', just as we should. And it is important to emphasise that these kind of syntactic transformations are not *ad hoc* additions to deal with substitution failures; it has always been understood that a system of syntactic transformations will be needed to take us from a surface-level sentence to its tree, and back again.[52]

Let's call the anonymous reviewer's version of substitution *tree-level substitution*. The first thing that the reviewer wanted to know was why we should prefer my sense-substitution to their tree-level substitution. However, I do not think that these two accounts of substitution are in competition at all: they work together, at different levels of abstraction. Imagine someone asked why the reviewer's account of tree-level substitution should make us say that 'I am hungry', the surface-level sentence itself, counts as a result of substituting 'I' for 'Robert' in 'Robert is hungry'. The answer would presumably be that this account shows that the surface difference between '( ) am hungry' and '( ) is hungry' is semantically insignificant: phrase structure trees are the primary objects of semantic evaluation, and the surface difference between '( ) am hungry' and '( ) is hungry' does not reflect any difference in the trees. But that is just another way of saying that these predicates have the same sense.

My sense-substitution is, then, just an abstraction from tree-level substitution. I skipped straight to the end of the story, which is what I take to be the most philosophically interesting bit. What we are left with is a general, and very flexible, notion of substitution. Of course, there is no sense in which a more abstract definition of substitution is automatically better or worse than a more concrete one. It all depends on the purposes you want to put it to. I would wager that sense-substitution is far too abstract for most purposes in linguistics; for those purposes we would be better off using something like tree-level substitution. But by contrast, I think that tree-level substitution is too concrete, too concerned with linguistic practicalities, for many philosophical purposes. At the very least, I hope that I managed to convince you that sense-substitution can be put to useful philosophical work over the course of this paper.

The anonymous reviewer had another question to ask about tree-level substitution. As I explained in Sect. 4.2, my definition of sense-substitution seems to run in a circle: I defined sense-substitution in terms of subsentential sameness of sense;

[52] For an excellent textbook introduction to phrase structure trees and natural language semantics, see: Heim and Kratzer (1998).

then I defined subsentential sameness of sense in terms of substitution; but it seems that the notion of substitution used in this last definition must be sense-substitution itself! The reviewer wanted to know whether we could break out of this circle by defining subsentential sameness of sense in terms of tree-level substitution. The answer is, I think, that we cannot, but it is worth taking the time to see why.

If we were only dealing with one language, *L*, then we could use something like tree-level substitution to define 'α has the same sense in *L* as β'. We start by moving from the surface-level of sentences in *L* to their phrase structure trees, and then say: α has the same sense in *L* as β just in case simple-substituting whatever corresponds to β for whatever corresponds to α never changes the interpretation of a tree, and *vice versa*. (Depending on what α and β are, they may correspond to lexical items, or they may correspond to subtrees.)

Importantly, however, this would not define a complex predicate, 'has the same sense in *L* as' with '*L*' as variable, but would instead define a simple predicate, 'has-the-same-sense-in-*L*-as'. That's because there is no one general method for moving from a surface-level sentence to its phrase structure tree, which could be applied neutrally to sentences from any language whatsoever. Different languages call for different transformations. So we would be left with a proliferation of synonymy relations, without the means to say what they have in common. Worse still, this whole approach goes no way towards explaining what it is for two expressions drawn from different languages to share a sense.

The trouble here is precisely that tree-level substitution is too concrete for a general definition of what it is for two expressions to have the same sense. We need an account of substitution which is more abstract, which steps back from the details of how any one language works. Of course, we have just such an account to hand, but it is sense-substitution itself. It seems to me, then, that the circle I described in Sect. 4.2 may well be unavoidable.

References

Bach, K. (1997). Do belief reports report beliefs? *Pacific Philosophical Quarterly*, *78*, 215–241.

Braithwaite, R. (Ed.). (1931). *The foundations of mathematics and other logical essays*. London: Routledge.

Dolby, D. (2009). The Reference Principle: A defence. *Analysis*, *69*, 286–296.

Frege, G. (1892). On sense and reference. In Beaney 1997 (pp. 151–171).

Frege, G. (1893). *Die Grundgesetze der Arithmetik* (Vol. I). Jena, Pohle.

Frege, G. (1897). Logic. In Beaney 1997 (pp. 227–250).

Frege, G. (1906). Foundations of geometry: Second series. In McGuinness 1984 (pp. 293–340).

Frege, G. (1918). Thought. In Beaney 1997 (pp. 325–45).

Frege, G. (1923). Compound thoughts. In McGuinness 1984 (pp. 390–406).

Hale, B., & Wright, C. (2001). *The reason's proper study*. Oxford: Oxford University Press.

Hale, B., & Wright, C. (2012). Horse sense. *The Journal of Philosophy*, *109*, 85–131.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.

Larson, R., & Segal, G. (1995). *Knowledge of meaning*. Cambridge: MIT Press.

Levine, J. (2013). Logic and solipsism. In Sullivan & Potter 2013 (pp. 170–238).

Lillehammer, H., & Mellor, D. (Eds.). (2005). *Ramsey's legacy*. Oxford: Oxford University Press.

MacBride, F. (2005). Ramsey on universals. In Lillehammer & Mellor 2005 (pp. 83–104).

MacBride, F. (2011). Impure reference: A way around the concept horse paradox. *Philosophical Perpectives*, *25*, 297–312.
Magidor, O. (2009). The last dogma of type confusions. *Proceedings of the Aristotelian Society*, *109*, 1–29.
McDowell, J. (1977). On the sense and reference of a proper name. *Mind*, *86*, 159–85.
McGuinness, B. (Ed.). (1984). *Collected papers on mathematics, logic and philosophy*. Oxford: Blackwell.
McKinsey, M. (1999). The semantics of belief ascriptions. *Noûs*, *33*, 519–557.
Moltmann, F. (2003a). Nominalizing quantifiers. *Journal of Philosophical Logic*, *32*, 445–81.
Moltmann, F. (2003b). Propositional attitudes without propositions. *Synthese*, *135*, 77–118.
Oliver, A. (2005). The reference principle. *Analysis*, *65*, 177–87.
Potter, M., & Ricketts, T. (Eds.). (2010). *The Cambridge companion to frege*. Cambridge: Cambridge University Press.
Prior, A. (1971). *Objects of thought*. Oxford: Oxford University Press.
Ramsey, F. (1925). Universals. In Braithwaite 1931 (pp. 112–137).
Rosefeldt, T. (2008). 'That'-clauses and non-nominal quantification. *Philosophical Studies*, *137*, 301–333.
Schiffer, S. (2003). *The things we mean*. Oxford: Oxford University Press.
Sullivan, P. (2010). Dummett's Frege. In Potter & Ricketts 2010 (pp. 86–117).
Sullivan, P., & Potter, M. (Eds.). (2013). *Wittgenstein's Tractatus*. Oxford: Oxford University Press.
Trueman, R. (2012a). Dolby substitution (where available). *Analysis, 72*, 98–102.
Trueman, R. (2012b). Neutralism within the semantic tradition. *Thought, 1,* 246–251.
Trueman, R. (2015). The concept horse with no name. *Philosophical Studies*, *172*, 1889–1906.
Trueman, R. (2017). The prenective view of propositional content. *Synthese*. doi:10.1007/s11229-016-1309-4.
Wright, C. (1998). Why Frege does not deserve his grain of salt. In Hale & Wright 2001 (pp. 72–90).