

Belief and Synonymy

Author(s): Tyler Burge

Source: *The Journal of Philosophy*, Vol. 75, No. 3 (Mar., 1978), pp. 119-138

Published by: Journal of Philosophy, Inc.

Stable URL: <http://www.jstor.org/stable/2025424>

Accessed: 11-04-2017 04:10 UTC

## REFERENCES

Linked references are available on JSTOR for this article:

[http://www.jstor.org/stable/2025424?seq=1&cid=pdf-reference#references\\_tab\\_contents](http://www.jstor.org/stable/2025424?seq=1&cid=pdf-reference#references_tab_contents)

You may need to log in to JSTOR to access the linked references.

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>



*Journal of Philosophy, Inc.* is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Philosophy*

---

---

# THE JOURNAL OF PHILOSOPHY

VOLUME LXXV, NO. 3, MARCH 1978

---

---

## BELIEF AND SYNONYMY \*

**T**HE question of what substitutions are admissible on purely logical grounds in oblique belief contexts presents no mere exercise in formal semantics. Dealing with it forces one to consider issues about the relation between language and mind, the community and the individual. The immediate aim of this paper is to argue that semantical theories that count exchanges of synonyms in belief contexts logically valid are inadequate as accounts of the notion of belief.<sup>1</sup> But this argument bears on larger issues regarding linguistic interpretation and the attribution of psychological states. In particular, it indicates specific grounds for rejecting the traditional model according to which the believer is directly and infallibly acquainted with the contents of his beliefs.

### I

In the 1950s, a remark by Benson Mates sparked a well-known debate over the conditions under which substitutions in belief contexts are logically guaranteed. The debate produced no consensus and was finally dropped. In this section and the next, I shall argue that Mates was more nearly right than his critics. In sections III–IV, I turn to the underlying issue.

Mates suggested taking distinct intensionally isomorphic or synonymous substituends for 'D' and 'D'' in the following schemas:

- (1) Whoever believes that *D*, believes that *D*
- (2) Whoever believes that *D*, believes that *D'*

\* I am grateful to Rogers Albritton, Tony Anderson, Ralph Kennedy, Richard Rorty, Andrew Woodfield for their suggestions. I have also benefited from discussions at Claremont College, California State University at Northridge, University of California at Riverside, and University of Wisconsin. My work was supported by the National Endowment for the Humanities.

<sup>1</sup> One may, if one likes, set up a logic of notions *like* belief which allows such exchanges—or even exchanges of logically equivalent expressions. But if these are not ordinary notions, some reason should be given for finding them interesting.

He then claimed that no one doubts that (1), and implied that someone might well doubt that (2).<sup>2</sup> The idea is that the possibility of a philosopher or other creature (we may call him "Bates") who, while believing that (1), is hesitant, on whatever grounds, to believe that (2), is sufficient to show that *D* and *D'* are not intersubstitutable in the belief context  $\ulcorner x$  believes that (1) $\urcorner$ . Bates's grounds need not be reasonable. As long as it is possible to believe that (1) while not believing that (2), the substitution cannot be justified as part of a *logic* governing belief sentences.

The two most detailed rejoinders to Mates's argument were those of Sellars and Church.<sup>3</sup> Sellars wrote that we must distinguish a "pure using use" and an autonymous (self-referential) use of expressions in oblique belief contexts. It is doubtful that expressions are ever *purely used* in oblique belief contexts. (See section III.) But there is a distinction between the occurrence of 'Hellenes' in

Jones believes that all Greeks are Hellenes.

where we understand 'Hellenes' in our sense, and 'Hellenes' in

The sentence "All Greeks are Hellenes" expresses something Jones believes.

where we assume that 'Hellenes' is to be understood as Jones uses it. We shall call the latter occurrence "direct discourse" occurrence.

Sellars claimed that Mates's exposition applied only to covertly direct-discourse occurrences of '*D*' and '*D*'. Thus Bates might well believe that

(3) For any  $\alpha$ , if  $\alpha$  believes "All Greeks are Greeks,"  $\alpha$  believes "All Greeks are Greeks."

but doubt that

(4) For any  $\alpha$ , if  $\alpha$  believes "All Greeks are Greeks,"  $\alpha$  believes "All Greeks are Hellenes."

where the quoted expressions are understood as  $\alpha$  uses them. For

<sup>2</sup> "Synonymity," in Leonard Linsky, ed., *Semantics and the Philosophy of Language* (Urbana: Univ. of Illinois Press, 1952), p. 215. Mates's argument was originally directed against Rudolf Carnap's theory in *Meaning and Necessity* (Chicago: University Press, 1956), where the notion of intensional isomorphism is defined. But the argument applies to most other formal theories of belief since Frege.

<sup>3</sup> Wilfrid Sellars, "Putnam on Synonymity and Belief," *Analysis*, xv, 5 (April 1955): 117-120. Alonzo Church, "Intensional Isomorphism and Identity of Belief," *Philosophical Studies*, v, 5 (October 1954): 65-73. For contrast, see Israel Scheffler's "On Synonymy and Indirect Discourse," *Philosophy of Science*, xxii, 1 (January 1955): 39-44.

example,  $\alpha$  might be a party to radically different conventions governing the relevant words.

The trouble with Sellars' rebuttal to Mates's argument is that it is too pat. Bates would certainly believe that (3) and doubt that (4)—almost anyone would. But this is not clearly relevant to the doubt Mates had in mind. Bates may not assume that the relevant  $\alpha$  uses the words 'Greeks' and 'Hellenes' at all. He may not have any particular counterexample to the substitution in mind. His doubt may be the result of a perfectly general caution about substitution in belief contexts.

Church made a slightly different appeal to direct-discourse belief. Using 'fortnight' and 'period of fourteen days' as the relevant synonymous expressions, he considers the following instances of Mates's (1) and (2):

- (5) Whoever believes that the seventh consulate of Marius lasted less than a fortnight, believes that the seventh consulate of Marius lasted less than a fortnight.
- (6) Whoever believes that the seventh consulate of Marius lasted less than a fortnight, believes that the seventh consulate of Marius lasted less than a period of fourteen days.

Church denies that anyone could possibly believe that (5) and doubt that (6), and claims that what Mates is really imagining Bates to doubt is that

- (7) Whoever satisfies in English the sentential matrix ' $x$  believes that the seventh consulate of Marius lasted less than a fortnight' satisfies in English the sentential matrix ' $x$  believes that the seventh consulate of Marius lasted less than a period of fourteen days'.

('Satisfies' is roughly the converse of 'is true of'; 'in English' assures that the quoted expression is construed as it is in English.)

This suggestion suffers from a defect analogous to that which defeated Sellars' suggestion: Mates need not have imagined an English-speaking skeptic. The difficulty can be repaired, however, by applying Church's strategy at a less embedded level. Thus Church could claim that what Mates himself believes is that someone (e.g., Bates) satisfies the matrix ' $x$  doubts . . .'. [Cf. (7).] Since this strategy is available and would not affect Church's argument for (7), I shall waive this objection.

Church argues for his position by appealing to the Langford translation test. First he translates (7) into German. When he sets out to translate (5) and (6), he notes that German has no single word that translates 'fortnight' and that the literal translation of

'fortnight' is the same as that for 'period of fourteen days'. He concludes from this that the translations of (5) and (6) are identical. Thus, Bates's supposedly different attitudes [belief that (5), doubt that (6)] could not be reported in German, at least not by making use of the usual translations. But this shows, Church believes, that the difference in attitude cannot be truly reported in English. On the other hand, a difference in attitude is translatable at the meta-linguistic level. And Church regards this as confirmation of his claim that Bates really doubts that (7) while believing that (5) and (6).

The main difficulty with the argument is that appeal to translation here is inappropriate.<sup>4</sup> Good translation should preserve truth value. But Church and Mates disagree about the truth value of

(8) Bates believes that (5) and does not believe (or doubts) that (6).

Church takes it to be a contradiction, whereas Mates regards it as true. Until this difference is resolved, one surely cannot decide whether Church's proposed translation of (8) (in the relevant context) as an *explicit contradiction* is correct. Obvious truths and falsehoods are normally used in arriving at a translation—not vice-versa.<sup>5</sup> In view of the relevant lexical difference between English and German and the fact that Mates intends his argument to turn partly on lexical considerations, determining a good translation seems more than normally dependent on determining the truth value of the English sentence.

## II

The original problem that Mates's argument presented to theories of substitution in belief contexts, like the theories of Frege, Church, and Carnap, was that it seemed to force defenders of those theories

<sup>4</sup> There are defects in the details of Church's argument which I shall skim over. There is an irrelevant focus on the question of whether a language has a *one-word* translation of a term. Moreover, if this focus is avoided, the *reductio* argument that supplements the appeal to translation begs the question. See Church, *op. cit.*, pp. 70/1. Note: Church stipulates that 'period of two weeks' is to be ignored as a possible synonym for 'fortnight'—similarly with its German translation.

<sup>5</sup> I regard translation as far less context-free than Church does. The assumption that 'fortnight' is to be translated 'Zeitraum von vierzehn Tagen' may well be overridden if more global considerations warrant. An adequate German translation in the present case would involve either ad hoc addition of 'fortnight' to German or a footnoted explanation of the point in question—no natural translation will do. The fact that a particular instance of Mates's argument cannot be naturally conveyed in German shows only that other instances would be better illustrations for German speakers. I discuss my view of translation and other aspects of the Langford test in "Self-Reference and Translation" in M. Guentner-Reutter and F. Guentner, eds., *Translation and Meaning* (London: Duckworth, 1977).

to hold that opponents like Bates or Mates either were believing analytical falsehoods or were not speaking English. Neither option is *prima facie* plausible. Church's reply to Mates appeared to offer an alternative: Bates and Mates could be construed as offering in English an empirical doubt about English, and taxed with a comparatively subtle use-mention confusion.

Despite appearances, however, Church's reply does not really provide such an alternative. This emerges as follows. Church holds that what Bates doubts is that (7). Interpreting Bates's doubt as metalinguistic will enable one to avoid attributing to Bates the original object-language doubt, only if one denies that Bates believes the most obvious truisms connecting metalinguistic semantical sentences with the object-language sentences. Ordinarily, we would assume that Bates believes that

- (9) For all  $y$ ,  $y$  satisfies in English the matrix ' $x$  believes that the seventh consulate of Marius lasted less than a fortnight' if and only if  $y$  believes that the seventh consulate of Marius lasted less than a fortnight.

and that

- (10) For all  $y$ ,  $y$  satisfies in English the matrix ' $x$  believes that the seventh consulate of Marius lasted less than a period of fourteen days' if and only if  $y$  believes that the seventh consulate of Marius lasted less than a period of fourteen days.

But if Bates doubts that (7), believes that (9), and believes that (10), he may reason to doubting that (6), if he has a minimum of logical acumen. For a little reflection indicates that the negation of (7), together with (9) and (10), *deductively* yields the negation of (6). But if we grant Bates the sense to believe that (5), we have reinstated Mates's original argument. The method of the argument we have just given is general in that it can be applied to any attempt to circumvent Mates's point by treating Bates's (or Mates's) attitudes as metalinguistic.

The argument seems to leave Church and other critics of Mates with two options. The first is to avoid reinterpretation or reconstruction and simply to say that their opponents are stating and believing analytical falsehoods. This alternative is not credible. In the first place, the intuitive attractiveness of Mates's point is left unaccounted for. In the second, the opponents are competent language users who understand on reflection the words they are using (they explicitly accept 'Greeks' and 'Hellenes' as synonyms); the assertions they make are neither particularly general nor particularly abstract; they remain unperturbed by the not especially

complicated counterarguments; and they give no sign of exceptional stupidity or irrationality. No one attributes analytical falsehoods under these conditions.<sup>6</sup>

The only remaining option for Church is to deny that Bates believes that (9) and that (10). What does such a response involve? The Tarskian biconditionals like (9) and (10), although contingent, express the most truistic of truths about English. To attribute a failure to believe such biconditionals (or what they express) is not to attribute an ordinary empirical doubt about English—nor is it to attribute a mere lack of theoretical understanding (unless the point of withholding attribution is simply that the subject has no semantical beliefs at all). For belief in such biconditionals, as Tarski realized, is among the nontheoretical, intuitive beliefs that a semantical theory must accord with. To attribute a failure to believe these biconditionals is to attribute inability to speak and understand the language in a quite nontheoretical sense of ‘understand’.

The response is preanalytically implausible. Bates’s philosophical or linguistic position (or caution about a position) does not seem to bespeak the elementary linguistic incompetence involved in such doubt. Ruling out Bates’s view, or his caution, by first translating his material attempts to express it into a metalinguistic mode, and then claiming that Bates is not competent as an English speaker would seem high-handed, and out of keeping with the initial plausibility of the view.<sup>7</sup>

I shall not try to state a criterion for believing an instance of the Tarskian biconditionals. I doubt that there is a precise criterion. Some general points, however, can be made. It is surely a necessary condition that one understand the language better than a foreigner does. A foreigner with only an understanding of ‘true’, quotes, and

<sup>6</sup> This claim raises difficult questions about disagreement over philosophical analyses. In my view, although synonymy has frequently been regarded as a model, philosophical analyses—even plausible ones—almost never resemble analytic equivalences. So attribution of erroneous philosophical analyses is not in fact attribution of analytical falsehoods. But even if it were, such disagreement occurs at very much higher levels of abstraction than the present dispute over Mates-like examples.

<sup>7</sup> The implausibility of the response can also be seen by shifting to indirect discourse. Virtually anyone would report Hilary Putnam in “Synonymy and the Analysis of Belief Sentences,” *Analysis*, xiv, 5 (March 1954): 114–122, as saying that someone doubts that whoever believes that all Greeks are Greeks, believes that all Greeks are Hellenes. But Putnam certainly did not say that someone doubts that whoever believes that all Greeks are Greeks, believes that all Greeks are Greeks. A German would give analogous reports. Note that the truth of Putnam’s claims is not at issue, but the truth of the reports of his claims. A Mates-like argument can thus be reproduced at any level of an infinite hierarchy.

'if and only if' might well believe of each instance of the truth schema that it is true. But this would not suffice to believe the instances themselves (have them as the contents of his beliefs). It is surely a sufficient condition for believing an instance that one have a general command of English together with competence in the use of the instantiated sentence (and 'true', quotes, and 'if and only if'). This competence might be demonstrated through ordinary mastery of the sentence's component words and its grammar. (It need not involve theoretical understanding of its logical form or its semantics, or even clarity about its verification conditions.) Clearly Mates, Burge, Bates *et alii* satisfy this test.

The main issue underlying the Mates controversy is when to re-interpret a compatriot's remarks in cases of apparent disagreement. I wish now to discuss that issue in greater depth.

### III

The examples that follow are *prima facie* incompatible with the view that synonyms are exchangeable *salva veritate* in belief contexts—even unembedded belief contexts.

I shall be using the word 'synonym' in an unloaded sense, as applying to terms so labeled in a dictionary. But the examples of synonyms I choose are the sort philosophers have appealed to in discussions of analytic truths. This last expression has no pre-philosophical sense; but I shall be (and have been) using it to apply to truths of quantification theory and arithmetic, and truths got from such truths by exchange of synonyms (in our unloaded sense). This is an attenuated usage, inasmuch as it does not carry the usual pretensions of "true by virtue of meaning alone, and with no thanks to the world"—a notion which I regard as empty since it lacks defensible contrast value. Thus my usage is an exhibition of bad philosophical taste, but hardly controversial.

We can imagine someone's saying,

- (11) I used to think that some Greeks are not Hellenes because I thought that Hellenes were only mainland Greeks, but now I know that all and only Greeks are Hellenes.

Of course, the person (if asked) would also say

I always believed that all and only Greeks are Greeks and never believed that some Greeks are not Greeks.

The example is not to be thought of as resting on use of purportedly synonymous proper names. Our points will apply equally to

- (12) I once believed that some female foxes are not vixens since I thought that a fox wasn't a vixen unless it had had sexual inter-



course, but I never thought that some female foxes are not female foxes.

and

- (13) For years I believed that a fortnight was ten days, not fourteen, though of course I never believed that fourteen days were ten days.

These quotations are not particularly startling or nonsensical. Ordinarily they would be taken at face value and as possibly being true. I myself could truly assert the last sentence. What they suggest is that sentences with forms like "All and only  $\phi$ s are  $\psi$ s" where distinct synonyms substitute for ' $\phi$ ' and ' $\psi$ ' may on occasion be used alone to give information or correct errors on the part of ordinary, generally competent language users. These points do not hold for sentences of the form "all  $\phi$ s are  $\phi$ s." If the examples are taken at face value, theories that allow substitutions of synonyms in belief contexts *on logical grounds* are wrong even for unembedded contexts. In what follows, I shall be assuming that it takes an *argument* to show that the examples should not be taken at face value. Other things being equal, literal (homophonic) interpretation within a language community is correct interpretation.

The view that examples like (11)–(13) should not be taken at face value is deeply ingrained in the analytic tradition. The demand for metalinguistic reinterpretation springs to the lips of any analytic philosopher so automatically as almost to preclude reflection. There are, however, several types of argument for such reinterpretation which can be elicited under pressure. I shall maintain that none of these provides good ground for reconstruing the sentences, giving them truth conditions other than those they appear to have, or regarding them as analytically or logically false.<sup>8</sup>

The first argument for not taking sentences like (11)–(13) at face value may be called the *i.e. argument*. This argument lies behind Sellars' attempted rebuttal of Mates: "In reporting someone to believe that some female foxes are not vixens, we presuppose that the words in our that-clause are to be interpreted as we use them in making normal assertions outside of belief contexts. Since we use 'vixen' in these latter contexts synonymously with 'female fox', we are saying that he believes that female foxes are not vixens, i.e., that

<sup>8</sup> The propositions that the speakers once believed *are* analytically false, but their statements and beliefs about their beliefs are not. I shall be using '*belief content*' to apply to the "expressive" (*de dicto*) part of whatever is denoted by that-clauses. '*Denote*' is a traditional technical term relating expressions to their referents in the context of a formal theory of truth.

female foxes are not female foxes. So substitutivity of synonyms holds. If the attribution of the mistaken belief is not to be reconstructed, it amounts to an attribution of an explicit contradiction.”

The premise of the argument is a half-truth and requires discussion. In effect, Frege denied the premise altogether, holding that words in that-clauses have senses other than those they have outside oblique contexts. Even apart from this viewpoint, the failures of the usual substitutions and generalizations suggest that expressions in that-clauses are not used in precisely the way they are in ordinary, transparent contexts. The cited points about presupposition do not support the conclusion. For it is only in explaining precisely how expressions in that-clauses are used (not in saying what is presupposed in using them) that one motivates the rules of substitution.

I shall not here try to justify my view of how expressions are used in that-clauses. My present argument does not depend on that view. But I will suggest, without elaboration, why I think Sellars' premise is half true. We do indeed presuppose that expressions occurring in that-clauses are to be interpreted as we (the reporters) use them outside of belief contexts. But, at least as a first approximation, expressions occurring in that-clauses function autonomously—to denote themselves. Our presupposed interpretation is of the denoted expression.<sup>9</sup> Rules of substitution such as Leibniz's law work on an expression's denotation, not its associated presuppositions. Taking expressions to denote themselves helps explain failures of exchange illustrated in (11)–(13). Certainly Sellars' argument provides no grounds against this viewpoint.

A second, far more common argument may be dubbed the *linguistic argument*. The argument focuses on the point that the speaker's original belief constituted a linguistic mistake, a mistake about language: “Since the speaker of (11)–(13) was guilty of a linguistic mistake at the relevant past time, his report of his past belief should be reformulated in the formal mode (with quotation marks), so that the belief was explicitly about language.” This claim is clearly part of the intuitive idea behind Church and Sellars' demand that Mates's doubt be reformulated as a metalinguistic doubt. But, as our argument in section II illustrated, the fact that a belief involves linguistic considerations does not show that the belief content must be formulated so as to contain quotation marks or to otherwise denote linguistic expressions. One's *reason* for giving up the belief that a fortnight is a period of ten days may in-

<sup>9</sup> I discuss this viewpoint at some length in “Self-Reference and Translation,” *op. cit.*, and in a sequel in preparation.

volve linguistic considerations, in the sense that the mistake was corrected by consulting a dictionary, or someone's linguistic intuitions. One might have come to realize that 'fortnight' means 14 days not 10. But that point does nothing toward showing that the relevant belief content in (13) must make reference to (denote) linguistic expressions. It is commonplace in logic to develop or discard beliefs expressed in an object language on the basis of metalinguistic reasoning. There is no reason why the same may not occur in everyday life.<sup>10</sup>

A related mistake in the linguistic argument is that of assuming that, since a question, statement, or belief is about language (in an ordinary sense of 'about'), it is used to denote a linguistic expression. The term 'about' is notoriously vague and notoriously not to be confused with 'denotes'. Sentences like "A fortnight is 14 days" are probably about language in the sense that the speaker's reasoning involved assumptions that denoted expressions, or in the sense that the normal means of confirming or disconfirming the speaker's beliefs would involve linguistic investigation, or in the sense that the speaker's reported beliefs are importantly equivalent with semantical statements. But none of these senses entails that the belief contents reported in (11)–(13) themselves denote linguistic expressions. Indeed, if there were such a straightforward inference from a statement's being about language to its being semantical (or metalinguistic), there would be no analytic truths of the sort got by exchange of synonyms. "All vixens are female foxes" is in an ordinary sense about language; its import is largely linguistic. But its logical form is not semantical.

Since the linguistic nature of the speaker's beliefs provides no basis for reinterpreting (11)–(13), one might hope to establish the reasonableness of reinterpretation by exploiting the kind of error the speaker committed. As we have seen, baldly asserting that the mistake was linguistic will not suffice. An argument made to the purpose is based on W. V. Quine's principle of charity, which in Quine's words runs as follows:

. . . assertions startlingly false on the face of them are likely to turn on hidden differences of language. The maxim is strong enough in all of us to swerve us even from the homophonic method that is so fundamental to the very acquisition and use of one's mother tongue. The

<sup>10</sup> It is not evident that, in thinking that some Greeks are not Hellenes, one *must* have *any* semantical beliefs (beliefs whose contents contain a semantical notion like truth or meaning). Primitive speakers tend to think at the object-language level, with no awareness of the conventional character of symbols. Cf. my "On Knowledge and Convention," *Philosophical Review*, LXXXIV, 2 (April 1975): 249–255.

common sense behind the maxim is that one's interlocutor's silliness beyond a certain point is less likely than bad translation—or in the domestic case, linguistic divergence.<sup>11</sup>

The argument is: "To attribute to an intelligent speaker the belief in an analytical falsehood (even for someone to attribute such to his past self) is uncharitable and not conducive to understanding or successful communication. Therefore we should not take his attributions at face value."

The idea behind the principle is to avoid attributing irrationality, particularly where it cannot be plausibly explained, if simple reconstrual of the subject's words would "rationalize" the situation. But a mistaken assertion or belief that a fortnight is a period of ten days would not ordinarily be irrational. Such a mistake would naturally be attributed to misplaced trust or slip of memory. To be sure, the mistake would be linguistic (in the sense discussed above) and would almost surely be correlated with and explained in terms of a mistaken belief that 'fortnight' in English means ten days. But granting that a belief that a fortnight is ten days stemmed from simple linguistic misinformation deprives one of grounds for holding that such a belief would be irrational. (One should resist the impulse here to revert to the linguistic argument.) Thus in the case as we are imagining it, Charity does not suffice to make us swerve from the homophonic method.

Arriving at a translation of a foreign language is relevantly different: attribution of belief in an analytic falsehood would not (ordinarily) be coupled with attribution of a *linguistic mistake*. The native understands his language better than we do. To avoid attributing inexplicable irrationality, we adjust our construal of the language. Even in the domestic case, we sometimes depart from homophonic construal. In cases of tongue slips, or of different dialects based on regional communal conventions, there is no linguistic mistake to "rationalize" an analytically false belief. Or in discourse involving vague or abstract expressions with no precise communally established usage (e.g., in discussing philosophy), we may avoid taking another's sentence in precisely the way we would use it, if doing so would rationalize his beliefs. But in the case we are imagining, belief in the analytic falsehood is fully compatible with rationality.

<sup>11</sup> *Word and Object* (Cambridge, Mass.: MIT Press, 1960), p. 59. I do not claim that the following argument would be endorsed by Quine, but it is suggested by much that he says. Cf. also Donald Davidson, "On Saying That," *Synthese*, xix, 1 (December 1968): 130–146, pp. 137/8. Davidson offers a mixture of the charity and speaker-meaning arguments; I discuss the latter next.

It is, I suppose, possible for someone to mislearn 'fourteen' in a way analogous to my mislearning of 'fortnight'. But given mundane facts about frequency of usage, it is improbable that a generally competent speaker would mislearn 'fourteen' yet understand 'days', 'is', and 'ten'. (I shall return to the notion of general competence.) So we are more inclined to expect that attribution of a belief that fourteen days is ten days betokens genuine arithmetic ineptitude, and is therefore to be undertaken with some trepidation. Thus, it is sometimes possible to attribute the previously cited belief and withhold the present one. There is no evident ground—either intuitive or pragmatic—to take attribution of the second to *follow logically* from attribution of the first. The principle of charity so used is abused.

A further weakness in the argument from charity is that communication and understanding are not furthered by reconstruing a speaker's words "A fortnight is a period of ten days." In the normal case, if someone were to say something with those words, he would be perfectly clear about how to confirm or disconfirm his statement or belief in the face of disagreement: he need only find a dictionary or someone whose linguistic intuitions both he and his opponent trust. (This fact also renders implausible any claim that the speaker's utterance or belief was simply a confusion and should receive no interpretation.) Of course, if the speaker refused to take the dictionary as authority and blandly (and credibly) explained that where he came from 'fortnight' is normally used to denote a period of ten days, *then* we would surely reconstrue his words. But in the normal case, reconstrual provides no special illumination. The point is not that avoiding reconstrual provides more illumination. It is rather that the speaker's claim can be just as well understood—and semantic ascent or consultation of a dictionary just as well initiated—if his words are taken at face value. And *ceteris paribus* the words should be so taken.

A final argument focuses on *speaker meaning*: "In expressing a belief with the words 'A fortnight is a period of ten days', the speaker must have meant something different by 'fortnight' than what we nondeviant English speakers mean by it. So it is mistaken to report him (even for him to report his past self) as having believed that a fortnight is a period of ten days." There is certainly a sense in which the premise of the argument is true: the speaker took 'fortnight' to mean "period of ten days" when he used it. He had the latter notion in mind. But there is a sense in which the premise ignores an important aspect of the speaker's attitude. The willingness of the speaker to submit his statement to the arbitration

of a dictionary indicates a commitment to having his words taken in their conventional sense, *whatever that sense is*.<sup>12</sup>

Which of these considerations takes precedence in determining how to interpret the speaker's statement and attitude in the present instance? The speaker's behavior, as we are imagining it, provides the key. His willingness to defer to a dictionary or the intuitions of other speakers and his and our unmixed feeling, after consulting the relevant authority, that he made a mistake suggest that the latter consideration is crucial.

The point takes shape when one considers possible reconstructions of the speaker's erstwhile utterance ('a fortnight is a period of ten days')—reconstructions that would enable us to avoid attributing statement of the relevant analytical falsehood. The most alluring reconstruction of the speaker's word 'fortnight' would be to take it to mean "period of ten days." In a sense, that is what the speaker meant by the word. But such reconstruction is inappropriate in the context of the sentences under (11)–(13). On such an interpretation the speaker would never have made a mistake in his *use* of the word 'fortnight'; for the utterance, 'A fortnight is a period of ten days', would be analytically true. Unconsciously idiosyncratic intentions are not an exculpation from error, but an explanation of it.

A second reconstruction involves holding that the speaker stated and believed an erroneous metalinguistic proposition about the use of 'fortnight' in English. Now the speaker probably did have an erroneous metalinguistic belief about 'fortnight' (cf. note 10). But that was not the belief he was stating in his utterance 'A fortnight is a period of ten days'. There are no *prima facie* grounds for taking the utterance as metalinguistic. And we can ensure that doing so would be out of keeping with the speaker's intentions by assuming that the speaker intended to state an object-language, analytic truth fully analogous to 'Every bachelor is unmarried'. In view of the failure of the linguistic and charity arguments, there is no reason for contorting the meaning of the utterance in a way contrary to the speaker's intentions.

At this point there may be a temptation to grant that the speaker stated or asserted that a fortnight is ten days, but insist that he believed only that 'fortnight' means ten days—not the object-language statement. Such a view is coherent; it is based on the assumption that what the speaker thinks a word means determines

<sup>12</sup> This pair of considerations is reminiscent of the distinction between referential and attributive uses of referring expressions. Cf. Keith Donnellan, "Reference and Definite Descriptions," *Philosophical Review*, LXXV, 2 (April 1966): 281–304.

what belief he can express in using it. The problem with the view is that it simply does not accord with ordinary attributions of belief. The attributions in (11)–(13) are clearly not metalinguistic: the copulas in (11)–(13) simply do not mean ‘means’. Nor does the assumption suggest a *reason* (either practical or theoretical) to change or reconstrue ordinary practice. (Cf. note 1.) When one asserts (11)–(13), one is fully aware that one used to think, e.g., that ‘fortnight’ meant ten days; yet one still attributes the object-language belief. No evident inconvenience, irrationality, or error ensues.

Assuming that (11)–(13) *are* to be taken at face value, the question arises how to interpret analogous cases that are not directly concerned with linguistic meaning. Considering these cases will bring out a certain open-endedness in our belief attributions, and certain *prima facie* differences between assertion and belief.

Suppose (i) that Alfred believes that a fortnight is a period of ten days; (ii) that Alfred sincerely says, ‘Bertrand will be gone precisely a fortnight’; and (iii) that Alfred believes that Bertrand will be gone for precisely ten days. We imagine that Alfred is by and large a normal, rational English speaker and intends to be conventionally understood in making the statement described in (ii). How should we interpret the situation?

I think there is little doubt that Alfred *said that* (and claimed that) Bertrand will be gone for precisely a fortnight. Nonphilosophical speakers seem to take this for granted.<sup>13</sup> Did Alfred *believe* that Bertrand would be gone a fortnight? The answer is much less clear-cut than in (11)–(13), where ordinary usage is definite. I have found nonphilosophical native speakers on both sides. But most lean negatively. An *affirmative* answer is typically based on a desire to maintain a close relation between sincere assertion and belief. This answer appears to be part of a wider emphasis on the role of belief in an account of the subject’s reasoning and explicit judgments. Since Alfred believes that a fortnight is a period of ten days and that Bertrand will be gone for ten days, it is only reasonable for him to conclude that Bertrand will be gone a fortnight. A *negative* answer typically rests upon a twofold emphasis on the connection between belief and understanding and on the role of belief in nonverbal behavior. Defenders of this view usually highlight our greater re-

<sup>13</sup> Suppose that (ii) obtained in the course of a bet. Alfred bets a pound on his statement (whose utterance is tape-recorded), and the House bets the same that what Alfred stated is false. Now imagine that Bertrand is gone for precisely ten days. Alfred demands that the House pay up. But after the ensuing discussion (complete with tape-recorder and dictionary), Alfred would come around. The bet would not be called off because of Alfred’s misunderstanding of ‘fortnight’; still less would Alfred get his money.

luctance to say that Alfred did not understand what it was that he believed than to say that he did not understand what it was that he asserted. They also tend to mention that Alfred had in mind a period of ten days when he uttered 'fortnight': Alfred's actions will be largely based on his belief that Bertrand will be gone for ten days. His linguistic mistake is irrelevant for such purposes as meeting the train.<sup>14</sup>

The two viewpoints probably represent conventionally divergent applications of 'believes', backed by divergent conceptions of the role of belief in explanation.<sup>15</sup> One or the other conception may be applied even by a single speaker, depending on the context of attribution. Both views are coherent. Each is capable of taking account of the other's emphases. The affirmative view can point out that nonverbal behavior is not all there is to behavior and that belief in what one does not fully understand is really pretty commonplace. Imagine, for example, an English major's belief that energy equals mass times the square of the speed of light. Half-learned expressions line the fringes of our vocabularies. We may allow such symbols to function in an account of someone's mental life even in cases where the person significantly misunderstands them [as in cases (11)–(13)]—as long as significant misunderstandings are circumscribed and relatively rare. On the other hand, the negative view can hold that Alfred's assertion was sincere in the sense that he thought it true, but that since Alfred misunderstood its content, his inner attitude or judgment was not directed toward what he said, but elsewhere. (Though abortive, the speaker-meaning argument was at least pregnant.)

*Both* viewpoints regarding Alfred's belief, together with assumptions argued for previously, are incompatible with the claim that

<sup>14</sup> There is a purely practical advantage for the negative view. In many contexts it would be misleading to announce that Alfred believed Bertrand would be gone a fortnight unless one added that Alfred believed that a fortnight is ten days. Avoiding the attribution obviates an ordinarily irrelevant rehearsal of Alfred's mundane miscue. I am sure that this explains much of the negative view's relative popularity. But I think it weighs little in adjudicating between the two views, given that both are held by a sizable number of nonphilosophical native speakers.

<sup>15</sup> A brief word about methodology. I do not aim for a "scientifically acceptable" reconstruction of the notion of belief which settles all cases or accords with some other philosophical ideal. My purpose is to use ordinary usage of and intuitions about belief sentences as the basis for an account of deductive inference patterns. Where nonphilosophical language users are significantly divided in their intuitions, I believe one learns more about a concept like belief by finding a rationale for the different conceptions than by rounding off the account in favor of one of the viewpoints. None of this is to deny that a formal theory of belief must meet standards of logical coherence.



exchange of synonyms in belief contexts is guaranteed on logical grounds. On the affirmative view, the point is obvious. We may assume that Alfred does not believe that Bertrand will be gone for a period of fourteen days. Attribution of explicitly contradictory arithmetical beliefs [assuming (iii)] seems neither natural nor explanatory. On the negative view, we may assume that Alfred believes that Bertrand will *not* be gone for a period of fourteen days. But it is implausible and explanatorily unhelpful to hold that Alfred believes that

(14) Bertrand will *not* be gone for a fortnight.

Alfred believes that a fortnight is a period of ten days and that Bertrand will be gone for ten days [(i) and (iii)]. To add that Alfred also believes that (14) would be to attribute an error in reasoning virtually as blatant as belief in an *explicit* contradiction. The linguistic error seems hardly so consequential. Despite the negative view's emphasis on nonverbal behavior, there is no reason why it should commit itself to implausibilities in the account of the subject's reasoning. It is more plausible, on this view, to avoid attributing the belief that (14) as well as the belief that Bertrand *will* be gone a fortnight.<sup>16</sup> But doing so is incompatible with the view that synonyms are exchangeable *salva veritate*. Thus the failures of substitutivity illustrated by (11)–(13) are contagious. They spread via normal assumptions about rationality to “nonlinguistic” contexts.

The burden of this section has so far been that ordinary language ill accords with post-Fregean theories of substitution. The counterexamples affect both assertion and belief and include both linguistic and nonlinguistic subject matters (although the root mistake by the believer was always linguistic). I have maintained that the arguments that have been stated or implied for reinterpreting the examples are failures.

Our discussion has brought out the domineering role of the presupposition that speakers are to be taken at their word; that is, literally or homophonically. The presupposition is sometimes voided, but rarely in cases where a speaker who is a normal member of the community, suffering no temporary lapse such as a tongue-slip, himself carries the presupposition. When the presupposition is in force, communal conventions about the meaning of a speaker's words tend

<sup>16</sup> Thus on the negative view, we withhold attribution of belief with terms (or relevantly exact translations) misunderstood by the believer, except in attributions like (11)–(13) which are the natural means of identifying his mistake. One might liberalize this policy by attributing beliefs in cases where the misunderstanding does not affect truth-value.

to override what a speaker mistakenly associates with his words in determining what he says and even, sometimes, believes.

An account of the conditions under which we do reinterpret ordinary discourse would require separate attention. But a tentative sketch is in order. It would be generally agreed that, apart from cases of temporary lapse by the speaker, departure from homophonic construal of familiar symbols is motivated on one of two grounds: incompetence on the part of the speaker, or his unwillingness to defer to local conventions. The important issue is how these grounds are spelled out.

The clearest cases of incompetence are those of a foreigner or a child who, though mouthing our words, does not have the slightest idea what they mean. In such cases we do not attribute assertions or beliefs on the basis of these words at all. We have seen, however, that localized incompetence on the part of a normally competent speaker sometimes does not deter homophonic (or literally translated) attributions of assertions or beliefs. It is important to distinguish cases here. Attribution of mistaken *object-language* "linguistic" assertions and beliefs, as in (11)–(13), seems to demand only that the subject have a general competence in the language or the relevant part of the language. Occasionally, such attribution may require even less. (The conditions are surely vague.) For example, we may perhaps make such attributions to someone who has partially learned a second language, if they do not conflict with attributions already justifiable by reference to the subject's native tongue. Attributions like those in (11)–(13) seem inapplicable to children, who lack a basic mastery of the language, regardless of how they seem to misapply the terms. For here, as in the case of a foreigner's chance mouthing of an analytically false sentence, we have poor material for attributing a linguistic belief at all. It is worth noting that although the examples in (11)–(13) involve "near misses," more radical mistakes may be naturally attributed ('An isotope is a certain geometric shape' or 'By profession, an ophthalmologist studies butterflies').

Attribution of ordinary "nonlinguistic" statements [like (14)] seems to be governed by similar conditions, at least for ordinary indirect discourse ("said that"). (Contrast Davidson, note 11.) With the use of verbs like 'claim' or 'maintain' our standards may be somewhat stricter, although it is arguable whether this is a matter of implicature or a matter affecting truth value. Homophonic attribution in these cases seems to require that the particular linguistic mistake be in some vague sense a "near miss" in "linguistic

space" and capable of misleading the audience. What this requirement amounts to probably varies with the term and the context of the report. Attribution of "*nonlinguistic*" beliefs [cf. (14)] requires at least this much, and for most people, even more: something like average competence in the use of the relevant terms.

The view about substitution of synonyms that I have been attacking is not without point. In reporting statements or beliefs, we are often indifferent about the exact words used in the report. One formulation is often as good as another, and normally we regard these different formulations as expressing what in the context amounts to the same thing. But our standards for judging formulations as expressing the same belief or statement vary from case to case. On occasion we draw the distinctions very finely [as in (11)–(13)]. Often we are grossly inclined, allowing even nonsynonymous exchanges. Our standards depend on the purposes of the report and on facts about the believer.

The attempt to account for the interchangeability of allowable reports in terms of logical rules obscures the context-dependence implicit in our usage. For we regard exchanges of synonymous terms in our attributions as expressing the same belief or statement only given certain premises about the background knowledge of the subject—including his linguistic mastery—and about the purposes of the report. I am not prepared to claim that no changes in wording are ever guaranteed on logical grounds alone. It may be that exchanges of clearly syncategorematic or abbreviatory synonyms, or certain purely grammatical transformations, may be justified. But in view of the examples like (11)–(13), a combination of extremely strict logical transformation rules together with context-dependent criteria for correct reporting seems inevitable in any plausible account of belief and indirect discourse. (Cf. note 1.) The denotations of that-clauses in *de dicto* propositional attitudes are, from a purely logical point of view, more fine-grained than ordinary linguistic meanings.

#### IV

The view that synonyms are logically exchangeable in belief contexts has usually been motivated, I think, by a certain assumption about the nature of assertion, belief, and other propositional attitudes. The assumption is most succinctly and radically stated by Russell:

Whenever a relation of supposing or judging occurs, the terms to which the supposing or judging mind is related by the relation of

supposing or judging must be terms with which the mind in question is acquainted . . . It seems to me that the truth of this principle is evident as soon as the principle is understood.

As is well known, Russell thought of acquaintance as a form of direct, infallible, nonperspectival knowledge. The principle, at least as applied to "terms" which are universals (such as concepts, ideas, attributes, meanings) is certainly not peculiar to Russell. Prior to the middle part of this century, the principle in one form or another was the dominant viewpoint in philosophy.<sup>17</sup> The picture originally behind the doctrine was that concepts, ideas, attributes, senses, or meanings are entities that occur in judgments "immediately before the mind," in something like the way sensations are supposed to, and that certain combinations of these entities are "self-evidently" true. The picture might be taken to put the relevant entities *in* individual minds, or it might place the entities in some objective realm immediately accessible to various minds.

This view has in recent years been decried by Wittgenstein, Quine, and others. But Russell's doctrine has retained its hold even on many who find the "mental eye" picture unattractive and Russell's vision of infallibility unpalatable. This vestigial influence has appeared most prominently in the tendency among analytic philosophers to reinterpret the words of compatriots far more readily than actual practice or rational considerations really warrant. This tendency emerges in the popularity of talk about idiolects. It appears in claims that meaning is a matter of conceptual role, or communicative intention, in the individual. It guides formulations of the problem that we cannot tell what a person believes unless we know what he means by his words, and we cannot tell what he means by his words unless we know what he believes (where 'means' is implicitly used in the individualistic, "has in mind" way; cf. note 11). In all these cases, the urge to reinterpret is motivated by a desire to make the content of the subject's speech or attitudes accord with what, either intuitively or by the interpreter's philosophical lights, the subject must *have in mind*. The linguistic,

<sup>17</sup> Russell, *Mysticism and Logic* (London: Allen & Unwin, 1959), p. 221. Plato, *Phaedrus*, 249b-c, *Phaedo* 74b6-c4. Descartes, *Philosophical Works*, 2 vols., Haldane and Ross, eds. (New York: Dover, 1955): *Rules for the Direction of the Mind*, sec. XII, vol. I, pp. 41-2, 45; *Principles of Philosophy*, Part I, xxxii-xxxv, vol. I, pp. 232-233; *Replies II*, vol. II, pp. 42, 52. Hume, *A Treatise of Human Nature*, I,3,5; II,2,6. Kant, *A Critique of Pure Reason*, A7 = B11. Frege, *The Foundations of Arithmetic*, section 105. G. E. Moore, *Principia Ethica*, section 86. A prominent exception to the dominant view is Leibniz, *Selections*, P. Wiener, ed. (New York: Scribner's, 1951), p. 286; also *Discourse on Metaphysics* VIII-IX.

charity, and speaker-meaning arguments of section III all carry this motivation.

The philosophical importance of (11)–(13), and the subsequent counterexamples to the view that synonyms are substitutable in belief contexts, is that they undermine even these innocent-seeming applications of Russell's doctrine. Our viewpoint suggests that the content of propositional attitudes, *even de dicto ones*, is not determinable merely by reference to individuals' streams of consciousness or unconsciousness. In our examples, the content of the belief has not been fully mastered by the believer. The relevant words in the believer's repertoire are clearly not sufficient to determine that content apart from construal of the words. For apart from construal, the words are mere types of shapes or sounds. The believer's own construal of the words does not determine the content. For if it did, the believer's belief, in the relevant cases, would be true. Rather, in these cases the content partly depends on linguistic conventions of a broader community.

The community has become prominent in recent theories of reference. What a speaker's term applies to is not purely a function of notions that he has mastered. Our argument is consonant with this line, but goes further. The appropriate interpretation of descriptive expressions in his statements, and even the content of his thought, are not purely a function of notions he has mastered.<sup>18</sup> The viewpoint opposes the traditional conception of the relation between meaning and belief. A person's beliefs cannot be simply characterized in terms of his direct acquaintance with meanings, or with concepts. Nor can the "self-evidence" of analytic truths be accepted without qualification. A rational believer may, despite reflection, *disbelieve* simple analytic truths obtained from explicit logical or arithmetic truths by exchanging synonyms.

TYLER BURGE

University of California, Los Angeles

<sup>18</sup> Putnam's emphasis on division of linguistic labor is akin to our view. Cf. "Reference and Meaning," this JOURNAL (1973), pp. 699–711, and "The Meaning of 'Meaning'" in *Mind, Language and Reality* (New York: Cambridge, 1975). His main thesis is, however, different. He argues that the *extensions* of certain general terms are fixed apart from notions speakers have mastered. Putnam then simply includes a word's extension in its meaning. Since his focus is different—with less attention to belief, synonymy, and partial understanding, and more concentration on necessity, reference, and acquiring a socially acceptable understanding—and since certain aspects of his interpretation of his examples are not fully clear to me, I am not sure how much ground is shared. Still, the general thrust of his argument is congenial. Cf. pp. 247 ff of the latter work.