

Quine and the Problem of Synonymy¹

Peter Pagin

1. Synonymy in the Two Dogmas framework

At the end of section three of *Two Dogmas*, Quine decides to turn his back on what he calls ‘the problem of synonymy’. By then he has spent two sections on discussing what synonymy might amount to, dismissing first *definition* and then *interchangeability salva veritate* as candidates for explaining it. In a sense, he does the right thing, for the discussion of synonymy in *Two Dogmas* is really a digression.

Quine’s main target, or one of his two main targets, is analyticity. But, as Tyler Burge (Burge 1992:4-10) and Paul Boghossian (Boghossian 1997:335-7), among others, have pointed out, there are several notions of analyticity in *Two Dogmas*. The first notion (1951:20) of an analytic truth is that of a truth grounded in meaning independently of matters of fact. The second (1951:23), is that of a truth that can be turned into a logical truth by putting synonyms for synonyms. The third and last notion (1951:43) is that of a truth that holds come what may, as opposed to a synthetic truth that holds contingently on experience. At no point does Quine comment on the relations between these different notions, and it is not easy to know what he thought.

The first and the third of these are different but closely related, the one concerning independence from facts and the other independence from experience. These notions are directly relevant to the holism Quine proposes in section six. They are relevant to his discussion of the nature of science, of the relation between theory and experience, and of that between theory and the world. The second analyticity notion, however, does not really belong in this framework. By the definition of the second notion, logical truths are unproblematically analytic. But Quine is explicit about rejecting the analytic/synthetic distinction, or the existence of analytic truths, when talking of the first notion—independence from facts—and of the third notion—independence from experience. So logical truths wouldn’t qualify as analytic in any of those two senses. From Quine’s own point of view, the second notion is not really compatible with the other two.²

But if that is so, what is, after all, the problem of synonymy? If Quine is right about analyticity in the first and third sense, then a logical truth like

(1) All unmarried men are unmarried

1. Earlier versions of this paper were presented at the conference *Fifty Years Without the Dogmas*, Humboldt University of Berlin, September 2001, and at the philosophy of language seminar at the department of philosophy, Stockholm University. I am grateful to comments from the participants on those occasions, and especially to Paul Boghossian, Kathrin Glüer and Åsa Wikforss.

2. More precisely, the following four sentences are jointly inconsistent: ‘all analytic sentences are true independently of matters of fact’, ‘no sentence is true independently of matters of fact’, ‘all logical truths are analytic sentences’, and ‘there are logical truths’. Analogously with ‘unrevisable’ instead of ‘true independently of matters of fact’.

is not true independently of matters of fact, nor true independently of experience. But then we will not get any such sentence either by interchanging synonyms in a logical truth. Specifically, on the assumption that ‘unmarried man’ is synonymous with ‘bachelor’, this will hold of

(2) All bachelors are unmarried.

(2) is hardly more independent of facts or experience than (1) is, and hence it isn’t analytic in any of the two central senses. Because of this, it does not really matter to Quine’s rejection of analyticity, in the first or third sense, whether we have an adequate definition of synonymy or not. Even if a definition can be devised that is respectable from a Quinean point of view, Quine’s criticism of the first and third notion remains intact. The literature has not always been clear on this point.

The synonymy discussion does, therefore, even though motivated by Carnap’s writings, stand out as a digression. It is, however, a digression of considerable independent interest.

2. The problem of synonymy in Two Dogmas

The problem of synonymy, as it is conceived in Two Dogmas, is that of explaining what synonymy is, without making use of any other notion that is as much in need of explanation as synonymy itself. So stated, however, the problem is not very well defined. Are we supposed to *analyze* the concept of synonymy (if such there is)? Or are we to provide a systematic *reconstruction* of our intuitive, pretheoretic conception of synonymy (if such there is)? Or are we to provide a reconstruction that as far as possible approximates the *extension* of our intuitive conception? I guess the last of these alternatives is most like what Quine had in mind. At least it is Quinean in spirit. Take what we have in the way of clear and respectable means of defining or explaining, and see if we can come up with a definition that at least approximately agrees in extension with what we are disposed, pretheoretically, to count as synonyms. It is plausible, I think, to see Quine as attempting this in sections two and three of Two Dogmas, but in fact, as we shall see shortly, Quine accepted deviations from the pretheoretic conception, thus making the problem easier.

In section two Quine correctly dismisses the appeal to definition. In section three he turns to interchangeability. In a nutshell, he reasons as follows: “First, we propose that two expressions are synonymous just if they can be interchanged *salva veritate* in all linguistic contexts, except where they are mentioned rather than used. But, if the language providing these contexts is extensional, then coextensive expressions meet this condition. Coextensiveness, however, falls short of (cognitive) synonymy. To compensate, we can either enrich the language by adding the adverb ‘necessarily’, or strengthen the condition, from interchangeability *salva veritate* to interchangeability *salva analyticitate*. These two options are in fact equivalent, for prefixing ‘necessarily’ to a sentence is just another way of saying that it is analytic. But now we have been moving in a circle, for an

explanation of analyticity is what we wanted in the first place". To this one can add, in the light of later sections, that the strengthening to interchangeability *salva analyticitate* doesn't do much good for synonymy if there aren't any analytic sentences to begin with.

Now, it doesn't require genius to notice that Quine has left many options unconsidered here. He hasn't considered alternative enrichments of the language, nor alternative interpretations of the modal adverb, even though he has discussed both at other places. In particular, however, he hasn't considered alternative properties to be preserved by an interchange.

We can in fact extract such a candidate property from Quine's own holistic picture of science, i.e. of total science as a fabric with theoretical sentences in the interior and with experience-confronting sentences at the periphery (1951:42-3). The relevant property is then, to use Quine's own metaphor, the *distance* to the periphery. That is, two expressions count as synonymous just if any substitution, in any use-context, of the one by the other results in a sentence having the same distance to the periphery as the original sentence. This is interchangeability *salva distantia*. As Quine himself explains it, (1951:43), the distance to the periphery consists in the relative likelihood of giving up the sentence in the face of recalcitrant experience. What is preserved, then, by the interchange of synonyms, is the subjects willingness to make revisions. For instance, if I hold 'brother' and 'male sibling' as synonymous, then I am as unwilling to give up

(3) Brothers are male siblings

as I am to give up

(4) Brothers are brothers.

I need not think that (4) is *immune* to revision, but I will be very unwilling to give up this and any other logical truth. I am then equally unwilling to give up sentences produced from logical truths by interchange of synonyms. This proposal has been worked out in greater detail elsewhere (Pagin 2001). Distance to the periphery is not a metric, but a position in a strict partial order of being *less revisable than*. The most revisable sentences are at the periphery, and the least revisable ones in the center.

If the language in question is extensional, then, given a few natural rationality constraints on revisions, the synonymy relation will be different from logical equivalence, but it will *not* be more strict than logical equivalence. That is, logically equivalent expressions count as synonymous. Because of that, this definition will deviate from intuitive synonymy judgments. This much deviation did not bother Quine, however, who wrote that

Statements may be said simply to be cognitively synonymous when their biconditional (the result of joining them by 'if and only if') is analytic (1951:32).

A biconditional whose immediate parts are logically equivalent is a logical truth, and

logical truths are, by the second notion, analytic. So Quine would accept logically equivalent sentences as synonymous. He also accepted disposition talk. Indeed, the *less-revisable-than* ordering relation is more or less formulated by Quine:

[...] in this relation of “germaneness” I envisage nothing more than a loose association reflecting the relative likelihood, in practice, of our choosing one statement rather than another for revision in the event of recalcitrant experience (1951:43).

The proposal therefore seems to be a solution to the problem of synonymy, as Quine presumably conceived it at the time of Two Dogmas.

3. The problem recast

Nonetheless, there is reason not to be satisfied with such a solution, and not only because of the remaining deviation from pretheoretic synonymy judgments. After all, synonymy is supposed to be sameness of *meaning*, or at least intralinguistic sameness of meaning. So there is reason to ask what the preservation of relative revisability has to do with meaning, and to answer this question we need to have at least some idea of what meaning might amount to.

Quine’s own preliminary suggestion of interchangeability *salva veritate* is not itself formulated in isolation from any idea of meaning. After all, why would interchangeability *salva veritate* be even a candidate for a synonymy relation? The answer, I guess, is twofold. First of all, there is an assumption that semantics is *compositional*: the meaning of a complex expression is determined by the meanings of its parts and its mode of composition. Simplifying a little bit, it follows, as Carnap remarked (1956:121-2), that a substitutivity, or interchangeability, principle holds: replace an expression by another expression with the same meaning, and the meaning of the whole remains the same. For two expressions to have the same meaning, i.e. to be synonymous, it is then necessary that interchanges in larger expressions—where they are used, not mentioned—are meaning preserving. The most conservative option is to make this a sufficient condition as well. Then the meaning of a single expression can be identified with the contribution it makes to the meaning of larger expressions containing it.

Secondly, there is an assumption that *truth* depends on meaning. Even if there are other determinants of truth value than facts and linguistic meaning—and we know that because of indexicality, at least, there are—two sentences with the same meaning, as potentially used by the same speaker at the same time, should have the same truth value. If again we take a conservative line on sentence meaning, we can identify it with *truth conditions*. For what Quine in *Word and Object* called ‘eternal sentences’ we can be even more conservative and identify sentence meaning with truth value. The result will be that two expressions are synonymous just in case they are interchangeable *salva veritate*. It is unlikely that this idea of synonymy could have anything going for it without these, or closely related, background assumptions about meaning.³

Still, there have been at least two reasons for preferring talk of synonymy at the expense of explicit talk of meaning. One is that the concept of meaning is badly defined and very difficult to characterize, and that we have reason to be sceptical about including meaning entities in our ontology. Quine exemplifies both attitudes already in *Two Dogmas* (1951:22). The constructive idea, expressed by Quine in several passages (e.g. 1960:201), is then that if we manage to make the synonymy relation precise, we can simply identify meanings with equivalence classes of synonymous expressions, since those would serve as counterparts to synonymy as well as anything else, and if synonymy can't be made precise, then we should reject meaning entities anyway.

But this option is not, I think, acceptable. First, if synonymy is understood as just *intralinguistic* sameness of meaning, as it usually is, then we get the odd result that two expressions belonging to different languages also differ in meaning, in virtue of that fact alone, regardless of similarity in usage. For their respective meanings will be different equivalence classes of expressions, each being a class of expressions of one language only. The relation of interlinguistic sameness of meaning will be well defined, but it will have an empty extension. Second, if we stick to the idea of compositionality, the meaning of the whole, and thus the truth value of a sentence, should depend on the meaning of the parts. But it is quite implausible that membership in a class of synonymous expressions is *that* very property of an expression that helps determine the truth value of a sentence containing it. Taking meanings to be equivalence classes of synonymous expressions is putting the cart before the horse; sameness of meaning should depend on meaning, not the other way around.

The second reason, however, challenges this order from considerations of determinacy. Given any intuitive idea of meaning properties, we run a risk—maybe a high risk—that the assignment of meaning properties to expressions remains indeterminate. Several different meaning distributions over the expressions of a language may be equally well supported by the facts and the principles that jointly determine what is right. By contrast, the intralinguistic synonymy relation, as governed e.g. by the idea of interchangeability, will be uniquely determined. Because of this, the objection concludes, we can make good sense of synonymy talk even if meaning is indeterminate. Therefore, we should stick to the former.

It is of course correct that substantial ideas of meaning properties and meaning assignment induce the risk of indeterminacy. And it is true that the intralinguistic synonymy relation can be made determinate even if the meaning assignment isn't. But the right conclusion isn't that we should skip talk of meaning. Rather, the right conclusion is that the dependence of synonymy on meaning must be generalized. What we should say isn't

3. The connection between synonymy, compositionality and truth is highlighted by the problem of belief sentences. Benson Mates (1950), later supported by Tyler Burge (1978), claimed that intuitively synonymous expressions need not be interchangeable *salva veritate* in belief contexts, and Mates took that to be a problem for the concept of synonymy. By contrast, Jeff Pelletier (1994), who also endorsed the claim of substitutivity failure, took it to be a problem, not for synonymy, but for the compositionality of natural language. In both cases, however, the connection between these concepts is reaffirmed.

simply that two expressions are synonymous just if they have *the* same meaning, but that they are synonymous just if *any* correct distribution of meanings over the language will assign the *same* meaning to both expressions. This definition indeed makes the synonymy relation determinate, even if meaning assignment itself is not (provided it *is* determinate whether or not a given meaning distribution is correct), and gives the simple definition as a special case if it is.

Given such an understanding of the relation between synonymy and meaning, is there a problem of synonymy, beyond the problem of knowing what meaning is? I think there is, and I think it is pretty much parallel to Quine's problem. For Quine, the problem was to match *pretheoretic* ideas of synonymy with what can be explained by clear and respectable means. We can phrase the remaining problem in similar terms: the problem is to match intuitive ideas of synonymy with our best (clear and respectable) account of meaning. Will our pretheoretic judgments of intralinguistic sameness and difference of meaning be supported by our best account of linguistic meaning?

What will be the nature of our best account? I think there is no good alternative here to following Quine, Donald Davidson and many others at least two steps on the way. First, meaning is language-independent. That is, there is no good sense to a notion of meaning which doesn't even allow expressions of different languages to be *compared* for sameness or difference in meaning. Second, what meaning an expression has is essentially connected with how speakers manage to communicate with the help of that expression. That is, a meaning property plays a role for what is communicated by means of expressions having that property. Even though the second point may be a little more controversial than the first, I don't think either is much disputed.

So far, there is not much to conclude about synonymy. Meaning properties may, for all that has been said so far, involve *more* than what is relevant for communication. There may be (perhaps small) differences in meaning that don't matter for communicative success. In that case synonymy, being the complete lack of difference in meaning, is not relevant.

However, at this point I prefer to follow Quine and Davidson, and Michael Dummett, a third step, parting company with others. I shall assume that what isn't relevant for success of communication isn't relevant for meaning either. Meaning properties *are* properties that play a certain role in an account of successful communication.⁴ That is, if the interchange of two expressions α and β cannot make a difference to communicative success in any possible context, then α and β are synonymous.

Given this assumption, we can recast the problem of synonymy as follows: does success or failure of communication ever depend on synonymy according to pretheoretic standards? If it doesn't, then there is no theoretic basis for pretheoretic synonymy. And then there is a problem of synonymy, i.e. pretheoretic synonymy.

4. Note that there is nothing behaviorist about this choice. I don't think that successful communication can be understood in behaviorist terms (cf. Pagin *unpublished*). Moreover, I do not assume here even that it is always knowable whether communication succeeds or not.

When turning to intersubjectivity for the basis of meaning, the first place to look is naturally Quine's account in *Word and Object*.

4. Synonymy and radical translation⁵

It is a common view, I think, that in *Word and Object* Quine took the critique of meaning, and thus of synonymy, one step further. It is not just that intuitive synonymy cannot be made clear sense of; according to our best approximation of interlinguistic meaning, there cannot be such a thing. That is the lesson of the indeterminacy of translation, according to this view.

This is a mistake, however. Since intralinguistic synonymy can survive indeterminacy of meaning, it can also survive indeterminacy of translation (the reason is given in the appendix). But the problem of synonymy doesn't end there. In a sense it only begins.

In *Word and Object*, Quine proposed the basic translation condition, for observation sentences, to be a matching of stimulus meanings. However, as stimulus meaning was defined (Quine 1960:32-33), in terms of a pair of classes of stimulations, it is not something that can be shared between speakers. For a stimulation is a stimulation of a person's nerve endings, and as long as two persons don't share nerve endings, they don't share stimulations. Quine discussed this in lectures in 1965, later published as 'Propositional objects' (Quine 1969). He there rejects the appeal to homology of receptors as a solution, 'not only because full homology is implausible, but because it surely ought not to matter' (Quine 1969:157). He briefly considers appeal both to a looser idea of resemblance of stimulations, and to an idea of similarity of 'barrages of outside forces' (Quine 1969:159), but ends the paper with speaking of the intersubjectivity of stimulus meaning as a problem.

The problem resurfaces more than twenty years later, at the Quine conference in St Louis 1988, with Davidson's 'Meaning, truth and evidence'. Davidson notes that if translation is to be based on similarity of stimulus meaning, then strange results are possible:

[...] let us imagine someone who, when a warthog trots by, has just the patterns of stimulation I have when there's a rabbit in view. Let us suppose the one-word sentence the warthog inspires him to assent to is 'Gavagai!'. Going by stimulus meaning, I translate his 'Gavagai!' by my 'Lo, a rabbit' though I see only a warthog and no rabbit when he says and believes (according to the proximal theory) that there is a rabbit (Davidson 1990:74).

Because of the possibility of such discrepancies, according to Davidson, Quine's account—i.e. what Davidson calls the 'proximal theory'—leads to skepticism (Davidson 1990:74). Either way, it clearly seems that in the circumstances imagined by Davidson, the right translation would be 'Warthog', not 'Rabbit'. What matters seems to be the

5. Sections 4 and 5 are digressive illustrations of the basic conflict. The main reasoning picks up with section 6.

shared, distal stimulus, not the unshareable proximal one.

This seems to have been Quine's own conclusion. In *Pursuit of Truth* he says:

The view that I have come to, regarding interpersonal likeness of stimulation, is rather that we can simply do without it. The observation sentence 'Rabbit' has its stimulus meaning for the linguist and 'Gavagai' has its for the native, but the affinity of the two sentences is to be sought in the externals of communication. The linguist notes the native's utterance of 'Gavagai' when he, in the native's position, might have said 'Rabbit'. So he tries bandying 'Gavagai' on occasions that would have prompted 'Rabbit', and looks to natives for approval. Encouraged, he tentatively adopts 'Rabbit' as translation (Quine 1992:42).

So, stimulus meaning is still there as the best respectable approximation to the intuitive idea of meaning, but translation isn't based on correlating instances of it.

The consequence of this choice is a certain tension between synonymy and the correctness of translations. The synonymy relation will be intrapersonal and based on internal facts, while the translation relation will be interpersonal and based on external facts.

To illustrate this, let's adapt an example of Carnap's (1956b:236-40). Carnap, in 'Meaning and synonymy in natural languages', was concerned with the possibility of empirical methods for determining the meaning a speaker attaches to a particular expression. For the extension of a particular predicate, within a surveyable region, the method consisted in simply eliciting a response from the speaker, as to whether he wanted to apply the predicate to a particular object or not. Carnap acknowledged some difficulties, because of the possibility of factual error on the speaker's part, and because of the general problem of induction in trying to project the total extension from the extension of a limited region.

An additional problem is introduced when it comes to determining intension. Carnap considers two hypotheses about what the German speaker Karl means by the word 'Pferd': on the one hand *horse*, and on the other *horse or unicorn*. Since unicorns don't exist, the two hypotheses determine the same extension, and they are therefore empirically equivalent in predicting the applications of the predicate by the speaker. A version of this example can illustrate the present difficulty. For assume that Karl has two words, 'Pferd' and 'Schmerd', and that the stimulus meanings are different: Karl is disposed to assent to 'Pferd' after stimulations typically caused by horses, and to assent to 'Schmerd' after stimulations typically caused by horses *as well as* to stimulations that would be caused by animals looking like unicorns. Because of this difference, the two words must be counted non-synonymous for Karl.

However, it is not clear that they should be differently translated, if the correctness of translation is determined by the speaker's dispositions with respect to distal stimuli. For with respect to distal stimuli, the speaker has the same disposition to apply both terms, the disposition to apply it to horses, and to nothing else, since there is nothing else to which he is disposed to apply either of them.

Carnap tries to solve his problem by indirect methods: we can describe unicorns to

Karl and ask whether he would apply his term to it, and we can show him a picture of a unicorn. The first method can work, of course, if we already have a well established translation manual for a large fragment for Karl's language, but if the problem afflicts all predicates, this cannot be assumed for a solution of the problem. The second method might be decisive as well, but it need not, for strictly speaking Karl is disposed to assent after stimulations from unicorns, not from pictures of unicorns. After all, my disposition to assent to 'unicorn' is not the same as my disposition to assent to 'picture of unicorn'. The picture method should work if it gave Karl the *illusion* of seeing a unicorn, but that is not so easily accomplished.

It is even more difficult than that, since for the sake of determining the correct translation between Karl's use of 'Pferd' or 'Schmerd', and George's use of the English 'horse', we would need to provide *both* speakers with an illusion of a unicorn, in order to see whether there would be any difference between their dispositions to assent to their respective words. But this is not enough either, for it is also required that the illusion is suitably *shared*. There must be a shared distal stimulus that would elicit the reactions of both speakers. It would not be enough to directly stimulate the nerve endings of each.

In this particular case I guess the normal intuition is that translating 'Schmerd' with 'horse' is simply wrong, and we would look for methods for finding this out without a direct inspection of proximal stimulations. But even if this is right, it is not so clear in the general case. For one speaker may have two terms, *a* and *b*, differing slightly in (proximal) stimulus meaning, even though it is practically impossible, or even nomically impossible, that any *distal* stimulus would cause a proximal stimulation of the one kind that is not also a proximal stimulation of the other kind. These terms are non-synonymous for the speaker. But with respect to *distal* stimulus meaning, they may both be translatable into the same expression *e* of another speaker, e.g. 'rabbit', and it is not obvious that either of the translations should be considered incorrect. The difference in stimulus meaning between *a* and *b* need not be relevant to the translation of them into another speaker's language. Regardless of what we should say in any particular case, there seems to be a general tension between the intrapersonal requirement of sameness of proximal stimulus meaning and the interpersonal requirement of sameness of distal stimulus meaning. If intralinguistic synonymy depends on the first, and *meaning* proper on the second, there is a genuine conflict.

5. A Fregean example

The conflict mentioned above may have the consequence that intralinguistic synonymy is a stricter relation than sameness of meaning. That is, two expressions may have the same meaning, but still not be intralinguistically synonymous. But if that can happen, what kind of relation is intralinguistic synonymy? You might think that this conflict is a result of rather specific ideas about stimulus meaning, or linguistic behaviorism in general. But this is not obviously the case.

There may, on the contrary, be a general tension between the need for intrapersonal

semantic distinctions and the need for interpersonal semantic sameness. Such a tension is exemplified in Frege’s ideas of sense. As Frege presented it, a linguistic expression expresses its sense, and the sense is what contains the mode of presentation of the referent. The mode of presentation of an object to a particular speaker corresponds, on common interpretations of Frege, to how the speaker *knows* or *thinks about* the referent. How the speaker thinks of an object may be more or less idiosyncratic, and Frege himself provides an example, in ‘Der Gedanke’ (1918:24-6), of two speakers communicating about a third person, presented to each of them in a different mode. The two speakers, Leo Peter and Herbert Garner, will then associate different propositions with the sentence

(5) Dr Gustav Lauben has been wounded

since they associate different senses with the proper name, even though the reference is the same. On Frege’s view, there is a communication failure, and he even claims that the two speakers don’t speak the same language (Frege 1918:25). In Frege’s example, it is a mere coincidence that the two ways of determining the reference yield the same result, but that feature can be changed. We may assume that because of e.g. the social context the two speakers are justified in believing that under the name ‘Dr Gustav Lauben’ they understand the same person, regardless of differences in mode of presentation. We may also assume that their respective modes of presentation are so related that it is impossible for them to pick out different referents given shared background knowledge of the speakers.⁶ In such a case, it is strongly counterintuitive to say that communication fails. If this intuition is right, then communication can succeed despite interpersonal differences in Fregean sense.

The introduction of sense is motivated by concerns of individual psychology. We need an account of differences for a single person in cognitive value between the sentences

(6) $a = a$
 $a = b$

and we need an account of substitution conditions of expressions in propositional attitude contexts. In both cases, possible mental states of a single person—knowing the one thing but not the other; believing the one thing but not the other—needs to be accounted for, and the concept of sense is the core of the account. Sameness of sense (for non-indexical expressions) is, if anything is, synonymy in Frege’s theory. The need for differences in sense goes beyond what is needed for communicative success.

This shows that the semantic tension between the intrapersonal and the interpersonal

6. For example, suppose that for Leo Peter the name ‘Dr Gustav Lauben’ is associated with the mode of presentation of being the physician who lives on the second floor, and that for Herbert Garner the name is associated with the mode of presentation of being the physician who delivered the second child of the emperor. Suppose, finally, that both know that the physician who lives on the second floor is the physician who delivered the second child of the emperor.

exists even in a more mentalistic setting. It does not show that there is problem with intralinguistic synonymy, partly because of the rather natural Fregean motivation for the distinctions between senses, and partly because we have considered only communicative success in a particular case. The tension will become acute only when we have distinctions between sense that *cannot ever* make a difference between communicative success and failure. But it seems the tension is in fact acute, for it seems that an account of pre-theoretic synonymy will have to appeal to features of the individual speaker that cannot matter to communicative success.

6. Synonymy and communication

First, what does it mean in general for intrapersonal synonymy to be relevant to communication? It cannot amount to the following idea, that synonymy between two sentences in the speaker's language guarantees that if communication with a hearer succeeds with the one sentence, then it succeeds with the other. This isn't the right idea, for synonymy, no matter how strict—as long as it can hold between different sentences—cannot guarantee that success is preserved by interchanges. The hearer may simply know the meaning of just one of the sentences.

Rather, I think it should be understood the following way. Whether a particular communicative event, involving a speaker, a hearer and an utterance, is successful depends on the hearer's response. The response can be e.g. that of thinking that it is a beautiful night, following an utterance of the sentence 'it is a beautiful night'. When the response is appropriate, e.g. because the hearer came to think just the same thought as the speaker expressed by the utterance, communication succeeds. Then, let's call two sentences *communicatively equivalent* just if the same hearer response, or responses, would be appropriate whichever of the two sentences would be used by the speaker. Intuitively, the same thing is communicated with the two sentences; that is, it *is* the same thing as far as communication is concerned.

Now we can say that synonymy is relevant for communication just if synonymy between sentences coincides with communicative equivalence. For subsentential expressions it must hold that they are synonymous just if a substitution of the one for the other produces a sentence synonymous with the original sentence. This is the adequate condition on synonymy, for if synonymy is a stricter relation than communicative equivalence, then two sentences need not be synonymous for being communicatively equivalent, and so the added similarity that synonymy contributes makes no difference to communication. If synonymy is a more inclusive relation, then different things are communicated with synonymous sentences, and so synonymy fails to represent what matters to communicative success.

What does matter to communicative success? Clearly, a difference in truth conditions between sentences can matter. It need not always matter, but if e.g. two predicates have different application conditions, there are at least some sentential contexts in which that difference is crucial to what is communicated. Suppose for a moment that truth condi-

tions, say in the sense of possible worlds truth conditions, is what in general matters to success. That is, if the hearer reacts by thinking a thought which is false in some possible circumstances where the speaker's thought is true, then communication has failed. By the present proposal of how synonymy matters to communication, this would mean that synonymy is relevant just in case synonymy between sentences coincides with identity of possible worlds truth conditions. However, it is common wisdom that synonymy is more narrow than identity of truth conditions. For instance, the two sentences

- (7) Pigs grunt
- (8) Pigs grunt and $2+2=4$

are true in the same possible worlds, but by pretheoretic standards, they are not synonymous. So if truth conditions are what matters to communication, then synonymy doesn't.

But surely more does matter to communication than possible worlds truth conditions, even if we abstract from all pragmatic aspects. For instance, it does make a difference whether I make you think that $2+2=4$ or that $16*7=112$, even though both hold in all possible worlds. The most natural way to represent this difference is to let thought content be individuated by semantic *structure*. Intuitively, the components of the thought that $2+2=4$ are different from the components of the thought that $16*7=112$ (except for the identity relation), and that is a reason for distinguishing between the thought contents as wholes, too.

The linguistic counterpart of structured thought content is structured meanings, and the classical definition of a synonymy relation based on semantic structure is Carnap's definition of *intensional isomorphism* (Carnap 1956:56-7). Briefly stated, for two expressions to be intensionally isomorphic, they must have the same syntactic structure, and corresponding *simple* constituents must have the same intension. If we apply these ideas and find that, first, sameness of structured thought content is necessary and sufficient for communicative success, and, second, that two sentences are communicatively equivalent just if they are intensionally isomorphic, and, thirdly, that intensional isomorphism coincides with synonymy, then we do have the result the intrapersonal synonymy is relevant for communication.

However, in so far as we are concerned with pretheoretic synonymy, the result is negative. Carnap himself notes that the expressions 'brother' and 'male sibling' are not synonymous by the standard of intensional isomorphism (1956:61), simply because one is simple and the other complex. But 'brother' and 'male sibling' is one of our stock examples of pretheoretic synonymy.⁷

It might also be the case that two simple expressions can have the same intension without being synonymous by pretheoretic standards, even though I don't know of any

7. In Pagin *forthcoming* I suggest a somewhat relaxed synonymy relation (dubbed 'μ-congruence') that allows a simple expression to be synonymous with a complex one, but otherwise is close to Carnap's. It is far from clear how that notion measures up to intuitions (it has been rejected as counterintuitive by a couple of linguists). The matter is discussed in the final section of Pagin *forthcoming*.

uncontroversial examples. One kind of example, although controversial, would be proper names, if one would combine the view that names are rigid designators with the view that names have Fregean senses. Given that they are rigid, coreferring names have the same possible worlds intension, but they can still differ in sense, and hence be non-synonymous.

Either way, it seems clear that pretheoretic synonymy does not coincide with intensional isomorphism. Neither does it seem easy to save this equivalence by some sophisticated syntactic theory. The idea would be that the criterion of intensional isomorphism should not be applied to surface structure, but to an underlying level of syntactic form. Some examples of apparent synonymy combined with surface structural difference can be plausibly handled this way. For instance, the sentences

(9) The police brought in the criminal

(10) The police brought the criminal in

would count as intuitively synonymous, and they can plausibly be seen as sharing an underlying structure.⁸ Applying the criterion at that level would make them come out isomorphic. Similarly, if the underlying form of ‘brother’ in fact is the same as that of ‘male sibling’, then the two expressions would count as intensionally isomorphic despite the surface difference. But I can see no good reason for thinking that general principles of syntactic theory would make *all* pretheoretical synonyms share an underlying structure. Rather, there is good reason to suspect that this could only be achieved by a series of *ad hoc* stipulations.

7. Beyond semantic structure

So far the result has been negative. However, even if we cannot represent pretheoretic synonymy by semantic structure, there must be some feature of the speaker’s linguistic dispositions that singles out synonymy pairs from other pairs, beyond his disposition to call the members ‘synonymous’. One such feature might be the sameness and differences between dispositions to revise judgments, which would be in accordance with the proposal in section 2. I am going to argue, however, that such a proposal does not meet the condition that meaning is to be relevant for the success of linguistic communication. In fact, the argument will be more general: *no* account of synonymy that goes beyond semantic structure will make the synonymy relation relevant for communication. The basic reason is that it is too implausible that the thought contents of speaker and hearer agree as far as semantic structure and that yet the communication fails.

8. (10) is derived from (9) by the separation transformation $T_{\text{set}}^{\text{op}}$ of Chomsky 1957, pp 75-76. In the days of transformational grammar it was argued, especially by Katz and Postal (1964:30-70), that all interpretation of complex expressions, i.e. by so-called projection rules, apply to *underlying* phrase structures. Thus, any singularly grammatical transformation, like from (9) to (10), is taken to be meaning preserving. In later theories, sentences are synonymous that share (interpreted) LF.

Assume that there is some relational property type *S* of expressions such that any two expressions sharing *S* type property are synonymous. The *S* type property may be what was suggested in section 2, i.e. what an expression contributes to the revisability conditions of sentences containing it. It can be some *normative* property, such as that of being stipulated to have a certain relevance for linguistic commitments etc. It can be a property that depends on the linguistic competence of the individual speaker, or a property that depends on general norms or regularities in the speech community. It will not matter which. All that matters is the assumption that two sentences can have different *S* properties, and hence be non-synonymous, even though sharing semantic structure, or have the same *S* property, and hence be synonymous, even though differing in semantic structure.

To consider an example, suppose that for speaker A

(11) x is to the left of y

is synonymous with

(12) y is to the right of x

for any uniform replacements of the variables. That is, an instance of (11) has a certain *S* property s^* for A iff the corresponding instance of (12) has s^* for A. Suppose, further, that this doesn't hold for speaker B. For B, any two corresponding instances of (11) and (12) will be necessarily equivalent, but they will not be synonymous.⁹

Now B says to A

(13) Elsa is to the left of Arnold

and we may suppose that communication is successful. A understands what B says. The question now is whether communication would have been successful if B had said instead

(14) Arnold is to the right of Elsa.

For B, (13) and (14) are equivalent (they will have the same truth value with respect to all possible circumstances), but they are not synonymous. Assuming that synonymy is relevant to communication, they are not communicatively equivalent, in the sense of the previous section. That is, the hearer response that is appropriate for a B utterance of (13) is not appropriate for a B utterance of (14). However, for A, the two sentences are synonymous, and hence the hearer response of A will, again on the assumption that synonymy is relevant to communication, be the same for both utterances. Then, since communication was, by assumption, successful when B used (13), it would have *failed*

9. If you have taken the *S* type to be a type that depends on what the speech community is like, you should here take A and B to come from different communities. These two communities can then be assumed to be completely alike except insofar as the *S* type properties, or some *S* type properties, are concerned.

had B instead used (14).

If the example is acceptable, two things are to be concluded. First, since the pretheoretic synonymy relation is allowed to vary between speakers, or between communities, pretheoretic synonymy in general is distinct from communicative equivalence. If A's synonymy relation does coincide with communicative equivalence, then B's doesn't (it makes too fine distinctions). If B's synonymy relation coincides with communicative equivalence, then A's doesn't (it makes too coarse distinctions). This holds regardless of how the communication in the example is evaluated.

Second, it is not reasonable to take the communication, by either (13) or (14) to have failed. We assumed that the S type properties go beyond truth conditions and semantic structure. Therefore, we can assume that as A and B understand the two sentences, they will have the same truth value in all possible circumstances. The same will hold for *any* pair of alternative instances of (11) and (12): communication between A and B cannot succeed with both members of the pair, even if A and B agree on possible worlds truth conditions of both members.

But this conclusion is too implausible. What more could be demanded for successful communication than that the meanings of all the parts of the sentence, and the structure of the sentence, uniformly contribute to determining truth conditions that will be the same as understood by both speakers?¹⁰ Setting the standards of success higher than that is simply a mistake. Such standards would deviate so much from common sense standards of communicative success that at best this would amount to a stipulation of a new meaning for the very expression 'communicative success'.

If this is right, then there cannot be any S type property, beyond semantic structure, that underlies synonymy. But maybe this verdict is premature. After all, if synonymy matters to communication, it is enough that the difference between A and B with respect to (11) and (12) *can* lead to communicative failure at some point when these expressions are involved. It should not be required that it lead to failure in the very instances of the schemata. But if not in the instances, where?

Maybe we should consider sentences like

(15) If Elsa is to the left of Arnold, then Arnold is to the right of Elsa.

For A, (15) would come out as analytic, in the sense of being transformable into a logical truth by replacing the occurrence of (14) with a second occurrence of (13). For B, since the two are not synonymous, (15) would be non-analytic in this sense. Should we not say, then, that communication between A and B by means of (15) would fail? Again, I think there is no good reason to do so, and the same good reason as before not to. Clas-

10. Clearly, (11) and (12) are semantically equivalent in the sense that every shared substitution (same term for same variable) produces instances with identical truth conditions. Therefore, it makes sense to count (13) and (14) as equivalent in semantic structure. Still, it is also possible to count them as not sharing semantic structure, e.g. since a direct application of the intensional isomorphism criterion gives the result that corresponding parts are not even coextensive.

sifying (15) as analytic for A and as non-analytic for B is simply a consequence of classifying (11) and (12) as synonymous for A and as non-synonymous for B. The former stands and falls with the latter, and does not add anything of its own.

Perhaps A is disposed to *call* (15) ‘analytic’, while B is not so disposed, but this does not show that there is any difference in how they interpret (15). All it shows is that either they understand (15) differently, or one of them makes a mistake, or they understand ‘analytic’ differently. Without a good reason for the first or second alternative, the third is the best choice. Similarly with calling (11) and (12) ‘synonymous’.

But couldn’t one say that for A (15) is a *conceptual* truth, while for B it is non-conceptual? I think there is no basis for this either. There is a good sense in which (15) is a conceptual truth for both A and B: the truth of (15) depends essentially on the semantics of the relational expressions, the conditional form, and the coreference of the singular term occurrences. As long as this is preserved, the singular terms can be varied *salva veritate*. For both A and B, then, the truth of (15) depends in this sense on the relational concepts, and not on the terms. But even so, it is left open whether (11) and (12) stand for, or express, the same or different concepts. Moreover, if they express different concepts, it is still left open whether the necessary equivalence depends on meaning alone, or if it depends on the nature of the corresponding properties (and it is left open whether this distinction makes sense).

All in all, (15) can’t make the difference. It should be clear that nothing would change in this respect by prefixing (15) with ‘necessarily’. But maybe belief contexts offer a decisive factor. For A, by assumption, would be disposed to assign the same truth value to

(16) Zeno believes that Elsa is to the left of Arnold

(17) Zeno believes that Arnold is to the right of Elsa

while it would be conceivable for B that one be true and the other false. So in this case, the difference would result in possibly divergent truth value assignments. Therefore, for B but not for A, there must be a difference in content between (16) and (17), and since this difference depends only on the difference between the embedded sentences (13) and (14), these must have a different content for B but not for A.

But this cannot work either. If synonymy makes a difference for belief contexts, then synonymy must already make a difference in other contexts as well. For if (16) is true and (17) false, then the beliefs expressed by (13) and (14) must *already* be different. If beliefs expressed by the two sentences are different, then the sentences are not communicatively equivalent to begin with, for it would be a misunderstanding to interpret an utterance of the one as expressing the belief that would be expressed by uttering the other. But if the beliefs are different, we want to know what it is that makes them different. The difference in belief content cannot *arise* only at the level of the belief sentence themselves. If B does hold that (16) and (17) differ in truth value, or that they could differ in truth value, there is a further question of what this attitude is based on.¹¹

The conclusion is that belief contexts do not offer a reason for making meaning distinctions beyond difference of semantic structure. There seems to be no good reason for the view that (11) and (12), for instance, could be synonymous for one speaker or speech community but just necessarily equivalent for another. If the communication condition is adequate, then there seems to be no good reason for assuming the existence of an S type property of the kind suggested.

Quine was perhaps wrong insofar as overrating the importance of the problem of synonymy to his overall theme of analyticity. He was wrong, I believe, in thinking that the problem of synonymy, as he conceived it in *Two Dogmas*, could not be solved. He was right, however, in pointing out the difficulty of saving or even approximating our pretheoretic notion, or notions, of synonymy by clear and respectable means. He was right, that is, in pointing out that there is a problem of synonymy. Moreover, his negative attitude to the prospects for solving it may in the end be borne out.

Department of philosophy
Stockholm University

Appendix: Synonymy and indeterminacy of translation

Quine's thesis of the indeterminacy of translation rules out non-linguistic meanings, such as propositions as the meanings of (eternal) declarative sentences, that can be shared by expressions of different languages. Quine himself is explicit about that at the end of §42. The reason is simple. The assumption to be reduced to absurdity is that correct translations are meaning preserving. Then suppose we have two correct translation manuals M_1 and M_2 between languages L and L^* . A translation manual is a definition of a recursive translation relation. By the indeterminacy thesis we can choose M_1 and M_2 so that the following holds: a sentence s of L is translated by M_1 into s' and by M_2 into s'' of L^* , but s' and s'' 'stand to each other in no plausible sort of equivalence however loose' (Quine 1960:27). So, s' and s'' do *not* have the same meaning. But, by the assumption of meaning preservation both s' and s'' have the same meaning as s . Given the transitivity of identity they also have the same meaning as one another, and so we have a contradiction. Hence, the assumption is false. Correct translations aren't meaning preserving. Even if there are propositions and other meanings, they are not determinately assigned to linguistic expressions.

If Quine is right, then, determinate interlinguistic sameness of meaning has to go, but there are still two questions about *intra*-linguistic synonymy. First, is the possibility of a well-defined synonymy relation compatible with the indeterminacy of translation? Second, if so, can it approximate our intuitive synonymy judgments?

Let's consider the first question. Our situation is pretty analogous to that in section 3,

11. Indeed, for Mates (1950) and Burge (1978), (16) and (17) could differ in truth value *even if* (13) and (14) are synonymous.

where we generalized the definition of synonymy to accommodate indeterminacy of meaning. The analogous solution would be this: we define two expressions e_1 and e_2 of language L to be synonymous just if it holds of any correct manual M, with L as the source language, that $M(e_1)=M(e_2)$. That is, they are mapped on the same expression by any correct manual M with L as the source language. Although this generalization accommodates the indeterminacy, it is not acceptable as it stands. For the target language may contain synonymous expressions, too. The simplest way of handling this complication is to take a manual to define in general a many-many relation and then to require the following: two expressions e_1 and e_2 of language L are synonymous just if it holds of any correct manual M, with L as the source language, and any expression e in the target language, that $M(e_1, e)$ iff $M(e_2, e)$ (that is, e_1 and e stand in the translation relation M just if e_2 and e do). Call this ‘translational synonymy’.

Note that this proposal does accommodate indeterminacy. To take Quine’s example of indeterminacy of reference, the term ‘gavagai’ of L can go into both ‘rabbit’ and ‘undetached rabbit part’ of English, by different manuals, but no single correct manual M would translate ‘gavagai’ with *both* of the English expressions. Thus, by the definition proposed, ‘rabbit’ and ‘undetached rabbit part’ are non-synonymous, as desired.

Could we then simply *define* intralinguistic synonymy this way? There are two problems with this suggestion. First, without a correct and complete set of criteria for the acceptability of a translation manual, we don’t have a complete definition of synonymy. We don’t know under what conditions of linguistic use two expressions of a language *should* be translated the same by any correct manual. Quine himself doesn’t propose any complete set, just a short list of desiderata, at the beginning of §15 of *Word and Object*. And in later writings he is even less specific (cf. 1993:48).

The second problem is that we don’t yet know whether what would be defined this way would be *synonymy*, or anything even close to it. It is pretty clear that we can claim two expressions to be synonymous if they are assigned the same *meaning* by any acceptable distribution of meanings over the language, but it is not so clear that we manage to approximate a pretheoretic understanding of synonymy by the condition that the two expressions be *translated* the same. We need to know more about the relation between meaning and translation. More precisely, under what relation between translation and meaning does the sameness-of-translation definition of synonymy imply the sameness-of-meaning definition?

To illustrate, assume that for any correct manual M, an expression e_1 of L is translated by M into expression e of L* just in case any meaning that can be correctly assigned to the one also can be correctly assigned to the other. Assume further that for both e_1 and e_2 of language L and for e of L* it holds that they can be assigned meanings m_1 and m_2 . Then, by the first assumption both e_1 and e_2 are translated by any correct manual M into e . In this way e_1 and e_2 can meet the conditions of the sameness-of-translation definition of synonymy. However, it is still possible that no *single* correct meaning distribution over L will assign e_1 and e_2 the same meaning. It may be that any total assignment will give m_1 to e_1 just if it gives m_2 to e_2 , and *vice versa*. So the conditions of the sameness-

of-meaning definition aren't met. The two expressions are translation synonymous but not (meaning) synonymous.

The situation changes, however, if the relation between meaning and translation is tightened. For suppose it holds that a manual M is correct just if any correct total distribution of meanings of the source language is also a correct total distribution over the target language under the correlation M . More precisely, for a manual M between L and L^* and a total distribution D of meanings over L , define an $M+D$ -induced distribution of meanings of L^* $M(D)$ so that $M(D)(e)=D(e_1)$ just in case $M(e_1, e)$. Then we can assume M to be correct just in case $M(D)$ is a correct distribution over L^* whenever D is a correct distribution over L .

Now the deviant translation example above is ruled out. For although there is, by assumption, a correct total distribution assigning m_1 to expression e of L^* , no total distribution assigns m_1 to both e_1 and e_2 of L . Therefore no correct manual M translates e with both e_1 and e_2 .

In fact, it is as easy to see that with this tightened condition on the relation between meaning and translation, translational synonymy does imply meaning synonymy. For suppose that e_1 and e_2 are translation synonymous. Then, by definition, any correct manual M such that $M(e_1, e)$ is also such that $M(e_2, e)$. Under the tightened relation between meaning and translation it holds that if D is a correct meaning distribution over L , then the induced distribution $M(D)$ is a correct meaning distribution over L^* . Therefore, it holds that $M(D)(e)=D(e_1)$ and $M(D)(e)=D(e_2)$, and hence that $D(e_1)=D(e_2)$. Since this holds for arbitrary correct M and D , the expressions are meaning synonymous if they are translation synonymous.

Whether this tightened relation really holds depends on the details of the correctness conditions of translations, which remain to be specified. Nonetheless, we can conclude that the indeterminacy of translation itself is no obstacle to having a well defined intra-linguistic synonymy relation.

References

- Boghossian, P.A., 1997, 'Analyticity', in Bob Hale and Crispin Wright (eds), *A Companion to the Philosophy of Language*, Blackwell, Oxford.
- Burge, T, 1978, 'Belief and synonymy', *Journal of Philosophy* 75, pp 119-38.
- Burge, T, 1992, 'Philosophy of language and mind 1950-1990', *The Philosophical Review* 101, pp 3-51.
- Carnap, R, 1956, *Meaning and Necessity*, second edition, The University of Chicago Press, Chicago.
- Carnap, R, 1956b, 'Meaning and synonymy in natural language', in Carnap 1956.
- Davidson, D, 1990, 'Meaning, truth and evidence', in R Barrett and R Gibson (eds), *Perspectives on Quine*, Basil Blackwell, Oxford.
- Frege, G, 1918, 'Der Gedanke. Eine logische Untersuchung', *Beiträge zur deutschen Idealismus* 2, 1918-19, pp 58-77. In English as 'The thought. A logical inquiry',

- translated by A.M. and Marcelle Quinton, in P F Strawson (ed), *Philosophical Logic*, Oxford University Press, Oxford 1967. Page references to the translation.
- Katz, J and Postal P M, 1964, *An Integrated Theory of Linguistic Descriptions*, MIT Press, Cambridge, Mass.
- Mates, B, 1950, 'Synonymity', *University of California Publications in Philosophy* 25:201-26. Reprinted in L Linsky (ed), *Semantics and the Philosophy of Language*, University of Illinois Press, Urbana 1952.
- Pagin, P, 2001, 'A Quinean definition of synonymy', *Erkenntnis* 55:7-32.
- Pagin, P, *forthcoming*, 'Communication and strong compositionality', forthcoming in *Journal of Philosophical Logic*.
- Pagin, P, *unpublished*, 'What is communicative success?'
- Pelletier, J, 1994, 'The principle of semantic compositionality', *Topoi* 13:11-24.
- Quine, W.V.O., 1951, 'Two dogmas of empiricism', *The Philosophical Review*.
Reprinted in reprinted in *From a Logical Point of View*, third edition, Harvard University Press, Cambridge, Mass. 1980. Page references to the reprint.
- Quine, W.V.O., 1960, *Word and Object*, MIT Press, Cambridge, Mass.
- Quine, W.V.O., 1969, 'Propositional objects', in Quine, *Ontological Relativity and Other Essays*, Columbia University Press, New York.
- Quine, W.V.O., 1992, *Pursuit of Truth*, 2nd ed., Harvard University Press, Cambridge, Mass.
- Quine, W.V.O., 1995, *From Stimulus to Science*, Harvard University Press, Cambridge, Mass.
- Salmon, N, 1986, *Frege's Puzzle*, MIT Press, Cambridge, Mass.