Fregeanism vs the Theory of Direct Reference

If I had to guess, here is how I would imagine that most people think of the philosophy of language after taking a standard undergraduate course. Pretty much every course will begin with Frege's distinction between sense and reference. Most students will find the distinction intuitive enough when they see the cases that Frege uses to motivate it. Some students will even be surprised that the instructor treats Frege's paper as a novel innovation since they already thought it was obvious that there must be some such distinction between different kinds of meaning. Later on, the students will read passages from the first two lectures of Kripke's *Naming and Necessity.* They will learn about Kripke's attacks on descriptivism, which are claimed to have undermined the Fregean conception of meaning. The reason that these arguments are supposed to undermine Frege's theory, they are told, is because Frege allegedly identified senses with descriptions grasped by the subject. But if this is all that they are told, then most students will be left with the impression that Frege was right in broad outline even if the descriptivist version of his theory is wrong in the details. After all, the Frege cases still appear to prove that there is a kind of meaning distinct from reference, even if it shouldn't be cashed out by descriptions.

For those who have this picture of the philosophy of language, no doubt it will come as a surprise to learn that most working philosophers of language no longer accept the Fregean two-tiered theory of meaning. (Some still do, but they're the minority.) Starting from the 1970s onwards, most philosophers have come around to accepting the "theory of direct reference", which says that the only kind of meaning possessed by a name is its referent. As the name suggests, the theory of direct reference claims that names refer *directly* to their object, without the mediation of a Fregean sense.

But isn't this regressive—a throwback to the time before Frege? Have these philosophers forgotten the *obvious* lesson of Frege's "On Sense and Reference"? Of course they haven't. However, it's hard not to get this impression if the considerations against Frege's solution haven't been sufficiently explained.

My aim for this post is to try to fill in some of this gap. That is, I want to summarize some of the reasons that philosophers have found to resist the Fregean conception of meaning. As always, since this is a blog post, none of this is intended to be terrifically ground-breaking or original. My purpose is only to clarify an issue that I frequently think requires clarification.

Fregeanism

Let's begin with Frege's theory of sense.[1] The best way to motivate the theory is to attend to the puzzles that give rise to the distinction between sense and reference.

Frege puzzles arise whenever a subject unwittingly uses two words to refer to the same thing, twice over, without realizing it. Famously, this allegedly happened to some ancient Babylonian astronomer who used the name "Hesperus" to refer to the first star to appear in the evening and the name "Phosphorus" to refer to the last star to appear in the morning. Little did they know that their terms "Hesperus" and "Phosphorus" referred to the same planet, Venus. It was only some time after the names were introduced that the Babylians found, through astronomical discovery, that the planet called "Hesperus" is the same as the planet called "Phosphorus."

A similar situation occurs to Lois Lane in the *Superman* comics. The person she knows as "Clark

[1] I have written a primer on the theory that can be found here: https://1000wordphilosophy.com/2020/11/19/freges-puzzle-and-the-meaning-of-words/

Kent" is the same as the person she knows as "Superman", although she does not know that they're the same. On a similar token, I have known some people who didn't realize that the philosopher named "Descartes" (who famously said "I think therefore I am") is the same as the Descartes who invented the Cartesian coordinate system and analytic geometry.

In each of these cases, there is a clear sense in which we want to say that both names in the pair share a common meaning: namely, they both *refer* to the same thing, and so it seems that reference is a kind of meaning for names. After all, the function of a name is to pick out its referent. When I say "Hesperus is a planet", I mean to be talking *about Hesperus.* What I have said is true or false depending on whether *Hesperus* is a planet.

If reference is all there is to the meaning of a name, then it follows that "Hesperus" means the same thing as "Phosphorus."[2] However, as Frege observed, there also appears to be a significant amount of evidence against identifying the meaning of a name with its referent alone.

For one, if meaning is simply reference, then the sentence "Hesperus is Phosphorus" will have the same meaning as "Hesperus is Hesperus." But as Frege pointed out, this appears to be wrong. The former sentence has the significance of an empirical discovery, whereas the latter is a mere tautology. Moreover, a competent subject (like the ancient Babylonian astronomer) can *understand* both sentences perfectly well, and yet they may doubt the truth of the former while affirming the latter. Taking different attitudes towards these two sentences does not make the astronomer irrational or linguistically incompetent or confused.

Frege took these observations to show that the names "Hesperus" and "Phosphorus" differ in what he called *cognitive significance.* Let's say that two names "e1" and "e2" differ in cognitive significance if it is possible for a competent, rational subject to grasp the meanings of "e1" and "e2" and take one attitude towards some sentence "e1 is F" while taking another attitude towards the sentence "e2 is F." Evidently "Hesperus" and "Phosphorus" differ in cognitive significance, and this is supposed to reveal that they differ in (some kind of) meaning.

We can strengthen the case by appealing to another assumption made by Frege. (We'll discuss this assumption momentarily.) Frege, along with most philosophers, have held that the meaning of a sentence is a *proposition.* But propositions are supposed to do double-duty as both the meanings of sentences and the contents of such mental states as *beliefs, desires, intentions, thoughts, etc.* If we assume that sentences have meanings that play both of those roles, and that the meaning of a sentence is determined by the meanings of the names that comprise it, then we get the result that the meaning of a name determines what thought (belief, desire, etc.) is expressed by a sentence. As a result, if we assume that the meaning of a name is its referent, then it would follow that *believing Hesperus is Phosphorus* is the same as *believing that Hesperus is Hesperus.* But again, that just seems plain wrong. It seems more right to say that the ancient astronomer *believed that Hesperus is Hesperus* but did not believe that *Hesperus is Phosphorus,* for the identity was not yet discovered. Similarly, it seems wrong to say that Lois Lane believed that Clark Kent is Superman prior to discovering his true identity.

Hence, Frege's puzzle. There appears to be some intuitive support for identifying the meaning of a name with its referent, and there also appears to be some intuitive support for distinguishing the meanings of some co-referential names. What do we do?

At this point I would like to caution the reader against making hasty conclusions based on their

[2] Consider this argument. "Hesperus" means *Hesperus*; "Phosphorus" means *Phosphorus*; and since *Hesperus is Phosphorus*, it follows that "Hesperus" *means the same thing* as "Phosphorus." Evidently we do speak of "*meaning*" in this way, and when we do, we are speaking of reference.

intuitive understanding of "meaning." Our intuitive understanding of meaning is highly vague. We understand meaning only through an open-ended list of platitudes about how meanings are supposed to function. For example, we think that (i) the meanings of words determines the meanings of sentences, (ii) understanding a language requires grasping the meaning of its words, (iii) the meaning of a sentence determines its truth conditions, (iv) words of different languages are intertranslatable on the basis of shared meaning, … etc.. There is probably a lot more to this list that anchors down our intuitive conception. Moreover, there's no guarantee that all of the items on such a list will be mutually satisfiable. Some of them may conflict with one another, in which case, we would need to refine our notion of meaning to make it consistent. This is the work for a *theory* of meaning. A philosophical theory of meaning will make explicit exactly which roles are to be performed by the meanings of a name or a sentence. The reason that Frege deserves credit, and why his paper is genuinely innovative, is because he actually makes it (relatively) clear exactly which roles the various meanings of an expression are supposed to perform.

The way that Frege proposes to solve the puzzle is by claiming that a name has not just one, but *two* levels (or kinds) of meaning. The first kind of meaning for a name is its ordinary referent. The second kind of meaning is what Frege calls "*sense*."

The basic idea is that, although "Hesperus" and "Phosphorus" share the same referent, they nonetheless express different senses. (So at one level they have the same "meaning", and at the other level they have different "meanings.") Moreover, the sense of an expression is supposed to explain the intuitive differences that we have witnessed above.

Throughout "On Sense and Reference", Frege offers various metaphors to explain the notion of sense. For instance, he says that the sense of an express is the "mode of presentation" of the referent. The idea here being that when Venus is called "Hesperus", it presents itself in a different mode than when it is called "Phosphorus", and these different modes are encoded into the different senses for the two terms. However, it must be admitted that this idea is fairly vague. For the core definition of sense, we must outline the roles that senses are supposed to perform. In particular, the sense of an expression is a feature that is supposed to fulfill the following jobs.[3]

(1) senses are a kind of meaning for linguistic expressions. Specifically, they are supposed to explain:

> (1a) *synonymy* (expresses are synonymous when they share the same sense);
> (1b) *linguistic understanding* (to understand an expression is to grasp its sense);
> (1c) the *publicity* of language (speakers share a common language when they associate its expressions with the same senses).

As a result, "Hesperus" and "Phosphorus" will *not* be synonyms because they express different senses.

(2) Senses are supposed to explain differences in *cognitive significance;* e.g. "Hesperus is Hesperus" is trivial while "Hesperus is Phosphorus" is informative *because* "Hesperus" and "Phosphorus" possess different senses.

(3) The sense of a name is supposed to determine its referent. This means that expressions with different

[3] Frege also appeals to sense to explain the significance of non-referring names. But this role is contentious and we will ignore it.

senses can share the same referent (e.g. “Hesperus” and “Phosphorus”), but expressions with the same sense must thereby share the same referent. Senses are thus meant to explain *why* an expression has its referent. (E.g. perhaps “Hesperus” refers to Hesperus because “Hesperus” expresses the sense <the first heavenly body to appear at night>).

(4) Senses are the constituents of thought. This explains why believing that *Hesperus is a planet* appears to be distinct from believing that *Phosphorus is a planet*; these beliefs *really are* distinct because they are composed of different senses.

(5) Senses serve as the referents of expressions within certain linguistic contexts, i.e. those that refer to meanings / propositions / thoughts. For example, when we utter “the Babylonians believed that Hesperus is not Phosphorus”, our expressions “Hesperus” and “Phosphorus” are referring to their ordinary senses, not their ordinary referents.

We can rigorously define a Fregean sense as any feature semantically encoded into an expression that satisfies 1 through 5. Notice that, now that these roles are laid out, Frege’s theory is no longer an obvious, unassailable item of common sense. It is a substantial theoretical conjecture that there is such a thing that fulfils all of roles 1 through 5, much less that roles 1 through 5 are mutually satisfiable. That isn’t simply given; it remains to be seen.

Why think that there is such a thing as sense? Well, we have already seen an argument based on the Hesperus/Phosphorus case. Having laid out the definition of sense, we can now make this argument more explicit.

The first premise of the argument is an empirical observation. Namely, that it is possible for someone (a subject, the astronomer) to understand both “Hesperus is Hesperus” and “Hesperus is Phosphorus” and rationally assent to the former while denying the latter, without linguistic confusion. From this, two things are supposed to follow: that the sentences have different meanings and that they express different thoughts. Now, if we take the meanings of sentences to be determined by the meanings of words, then it follows that “Hesperus” and “Phosphorus” must have different meanings and express different constituents of thought. This is despite the fact that they share the same referent. So it follows that there must be another kind of meaning, distinct from reference, that fulfills roles 2 and 4.

Moreover, since senses are still supposed to be a genuine kind of *meaning*, it only makes sense to attribute to them the roles outlined in 1. Any kind of meaning worthy of the name must explain *synonymy, linguistic understanding*, and the *intersubjectivity* of meaning. (Although, we will see later how this causes some internal tension within the concept of Fregean sense.)

Finally, what of roles 3 and 5? We’ll see shortly that 5 is justified by the orthodox theory of propositions which was partly pioneered by Frege. Role 3, on the other hand, can be given an intuitive gloss. If two, e1 and e2, *refer* to different things, then they cannot have the same meaning (in the same context). Why not? Well, one reason is that meaning is supposed to determine whether a sentence is true or false, along with the facts of the world. If e1 and e2 have different referents, then the sentences “e1 is F” and “e2 is F” may differ in truth value. In which case, they had better not have the same meaning. (There is another philosophical motivation for thinking of senses as reference fixers; I’ll explain it momentarily.)

The above argument for the existence of senses relies on two principles that we ought to make explicit. The first principle is one that concerns the compositionality of meaning:

(*Compositionality*) The meaning of a sentence is determined by the meanings of its parts (including names contained therein).

Admittedly, this principle is still vague as it stands, and there is a great deal of work that needs to be done to make it more precise. But regardless, it is supposed to constrain any kind of meaning, including both sense and reference. It is worth mentioning at this point that Frege conceived of both sense and reference as kinds of meaning that can be ascribed to other expressions besides names. Predicates also have reference and sense, as do whole sentences. We'll concern ourselves with sentences in the next section. Suffice to say that it takes a bit more subtlety to draw the distinction for other grammatical categories. We will save ourselves a lot of time if we only focus on names.

The second principle is what Gareth Evans calls the "intuitive principle of difference." Basically, this principle is needed to take us from our observations about speaker behaviour to a conclusion about meaning. In full generality, it states:

(*Frege's Principle)* If a competent speaker who understands sentences *S1* and *S2* can rationally assent to *S1* while dissenting from *S2*, then *S1* and *S2* express different meanings.

This is the principle that really drives the argument for Fregean senses. Intuitively, the idea is that meanings, *whatever they are*, ought to coincide (and explain) rational speaker behaviour. Hence, wherever a rational speaker can draw a distinction, we ought to recognize a difference in meaning.

Fregean Propositions

Apart from the Fregean conception of meaning for names, we should also consider its conception of *propositions.* But before we can do that, let's first say a bit about propositions. Propositions are a class of theoretical posits that are introduced by philosophers to fulfill five distinct roles. (Once again, a theory of meaning is made explicit by listing the roles that meanings are meant to serve.) To be specific, propositions are defined as doing these five things:

(i) they are primary bearers of truth and falsity;
(ii) they are the primary operands of the modal operators (it is necessary that…, it is possible that…, it is contingent that…, etc.)
(iii) they are the *meanings of sentences*. When a sentence is uttered in context, it will *express a proposition*. The proposition that it expresses will *encode its meaning*. Moreover, other sentences (from other languages or other contexts) will share the same meaning *in virtue of expressing the same proposition*.
(iv) Propositions are supposed to serve as the objects of the so-called propositional attitudes. They are the things that we believe, that we desire, that we intend, and so on. When I believe that it is raining, then I bear an attitude (belief) towards a proposition, *that it is raining*.
(v) Finally propositions are supposed to be the referents of "that" clauses. *"That it is raining"* functions like a referring term that refers to a certain proposition.

It may be disputable whether there really are such things that play all five of these roles. (To name one potential problem, I think that it is a substantial assumption that the meanings of the sentences of public language are one and the same as the objects of belief and desire. It seems to me to be an open possibility that public language encodes less fine-grained information than our mental states.) But the orthodox view of propositions, inspired by Frege, holds that sentence-meaning and psychological content are one and the same. We will adopt this assumption for the sake of discussing the theory. (If we were to reject this assumption, then we would have to discuss two distinct Frege puzzles: one concerning language and one concerning psychological states.)

With the concept of a *proposition* at hand, we can now characterize the Fregean conception of propositions. For Frege and his followers, the proposition expressed by a sentence is the *sense* of a sentence. Moreover, a proposition is composed of the senses of the sentence's constituents. This gives us an answer as to what a name, used in a sentence, contributes to the proposition expressed by that sentence. According to the Fregean theory, a name contributes its *sense*.

The theory of direct reference

Let's now contrast the Fregean theory of meaning with its primary competitor, which is sometimes called the *naive theory of reference, the theory of direct reference,* and *Millianism.* Whatever it's called, the main competitor to Frege's theory holds that the only kind of meaning possessed by a name is its *referent.* Thus the meaning of "Hesperus" is *Hesperus*, a.ka. Venus. The planet Venus itself serves as the meaning. Likewise, the meaning of "Phosphorus" is also the planet Venus itself. The reference between a name and its object is "direct" because it is not mediated by a Fregean sense.

One point may require clarification. In saying that the "Hesperus" and "Phosphorus" have the same meaning, the direct reference theorist is *not* claiming that the two expressions are equivalent in all of their features that are significant for language use. The theory of direct reference allows that these expressions may be used differently by a speaker, have different "connotations", or be tied to different conceptions within the mind of speakers. For all that the theory of direct reference claims, there may be fairly significant differences between the two expressions. Their only claim is that, whatever differences they have, they do not amount to differences in *meaning.* Or to put it another way, our theory of meaning need not pay attention to whatever differences there are between these expressions. The differences are real, but they aren't *semantic* differences.

(What makes a difference a *semantic* difference?, you may ask. Recall what I said earlier. Our intuitive conception of meaning is vague. It is loosely defined by various roles that meanings allegedly play, but perhaps not all of them will make the final cut in a well-developed theory. A theory of semantic content is, in part, in the business of sorting out which roles are essential to meanings and which ought to be played by some other feature of language use.)

The theory of direct reference goes hand in hand with its own conception of propositions. This is often called the Russellian view of propositions, after Bertrand Russell, since he pioneered the metaphysics, even though he rejected the theory of direct reference for names. According to this theory, a name, used in a sentence, contributes only its referent to the proposition expressed by that sentence. It follows then that propositions are composed of worldly objects and properties (the meanings of predicates, quantifiers, etc.). Thus according to this theory, propositions are pictured as literally containing objects from out there in the world.

Take the sentence "Socrates is wise." According to the direct reference theorist, this proposition will be composed of *Socrates* (himself) and *the property of wisdom* (itself). But for the Fregean, the proposition will not be composed of Socrates himself. Rather, it will be composed of the sense of "Socrates", i.e. the *mode of presentation of Socrates.* Thus the entire proposition will be a complex of *the mode of presentation of Socrates* and *the mode of presentation of wisdom.*

Of course, talk about what propositions are "made of" is a little spooky. The issue should not be thought of as primarily about the ingredients that figure into the construction of propositions. Instead, what's really at issue is the *individuation* of propositions. The questions that matter are really about whether certain pairs of sentences have the same meaning, or whether certain pairs of beliefs (or desires, etc.) are to be classified as instances of the same mental state. Again, for the Fregean, "Hesperus is a planet" has a different meaning from "Phosphorus is a planet", and believing that *Hesperus is a planet* is different from believing that *Phosphorus is a planet.* But for the direct reference theorist, the two sentences have the same meaning and the two beliefs have the same content.

Fregean sense in practice

So far I have been fairly silent on the entities that serve as Fregean senses. I have defined them by roles 1 through 5, but that is only to say what they *do*—it is not yet to say what they are. The question of identifying Fregean senses beyond the roles that they're intended to play turns out to be incredibly vexed. I'm not going to get too much into it, except to make one point.

By far the most common way to think of Fregean senses is to think of them as *concepts* that are grasped by the speaker. (However this doesn't make things terrifically more clear. What is a concept, anyway?) Thinking of senses as concepts suggests that they encode the speaker's *conception* of the referent. As such, they encode information that the speaker has about the referent.

This account of Fregean sense gives rise to the descriptivist interpretation, whereby the sense of an expression is a description in the mind of a speaker. Frege himself gives examples that suggest this interpretation. On this account, the sense of "Hesperus" would be something like <the first heavenly body visible in the evening> and the sense of "Phosphorus" would be <the last heavenly body visible in the morning>.

This descriptivist interpretation does a good job of explaining why senses have the roles that are ascribed to them. For instance, it makes clear how Fregean senses can play the role of reference-fixers (role 3), whereas other interpretations will leave this role opaque. The idea here is that an expression, like "Socrates" will express a sense, such as <the snubbed-nosed ancient Greek philosopher who was tried and executed for corrupting the youth>, which is grasped by speakers who are familiar with the expression. This description determines that it must be *Socrates*, in particular, who is the referent of the term, in virtue of him being the unique satisfier of the conditions encoded in the description.

The descriptivist interpretation is also tailor-made to explain the epistemic role of senses outlined in role 2. Suppose that "Hesperus is Hesperus" expresses the proposition <the first heavenly body visible in the evening is the first heavenly body visible in the evening> and "Hesperus is Phosphorus" expresses the proposition <the first heavenly body visible in the evening is the last heavenly body visible in the morning>. Plainly, this gives an explanation of the epistemic differences between the two sentences. The first proposition is knowable a priori whereas the second one is only knowable a posteriori.

The idea that the meaning of an expression encodes information that must be satisfied by its referent has an esteemed history in Western philosophy owing to its ties to the program of conceptual analysis. There was a time when analytic philosophy conceived itself as largely in the business of discovering conceptual truths by reflecting on the meanings of certain key terms. The truths thus discovered would be deemed *analytic*. Since they were supposed to be incorrigible and knowable *a priori*, they were held to have a special role in our body of knowledge.

This conception of philosophical methodology, along with its robust notion of analyticity, goes hand-in-hand with the Fregean conception of meaning. For, according to the Fregean account, words express complexes of information that are grasped by competent speakers. This allows for simple words to bear logical connections to each other which are knowable simply by virtue of linguistic competence.

To give an old (and outdated) example, it used to be thought that knowledge is simply justified true belief. How could one (apparently) know this (alleged) fact? One answer, typical of the Fregean, is that the word "knowledge" expresses a sense which encodes these three conditions <*justified, true, belief*>. This is claimed to be known by anyone who understands the word "knowledge" and grasps this sense. That *knowledge is justified true belief* would thus be construed as tantamount to the proposition that <justified true belief is justified true belief> , a logical tautology. Moreover, according to the Fregean conception, the word "knowledge" would only refer to a property *if* that property met the conditions that are encoded in its sense. Thus a subject does not require familiarity with the *referent* of "knowledge" in order to know that it is justified true belief; they only need familiarity with the *sense* of the word, and the sense *guarantees* that the referent meets those three conditions. In other words, mere linguistic knowledge is sufficient to "know" this alleged analytic truth.

Nowadays, nobody can get away with claiming that knowledge of analytic truths is this easy to come by. Philosophically significant examples of analytic truths are very hard to produce, if not impossible. (The account of knowledge as justified true belief was discredited by Gettier in the 1960s.) Indeed, the fact that there are so few interesting analytic truths is something of a disappointment for (this version of) Frege's theory of meaning. It is also a mark against the view. Competent speakers can disagree over whether a given sentence expresses an analytic truth, without forfeiting their linguistic competency. This undermines the picture of meaning that sees it as encoding robust information about the referent, graspable by all competent speakers.

Meaning vs Reference Fixer

We can now finally turn to the considerations that undermine the Fregean conception of meaning. Here is how I suggest that we understand the dialectic. First of all, there is an open-ended list of pre-theoretic desiderata for theory of meaning. For instance, we think that meaning explains the determination of truth conditions, has a role in explaining linguistic competence, that the meanings of complex expressions are determined by the meanings of their constituents, and so on. Frege's theory offers us a specific list of the roles that are supposed to be definitive of one kind of meaning. To test whether it is a good theory, we must see whether the roles are mutually satisfiable by a single feature encoded into expressions. The arguments against Frege's theory essentially aim to undermine the idea that there is any one feature that plays all of the roles attributable to sense.

Let's start with the classic arguments for content externalism from Kripke's *Naming and Necessity* and Putnam's "Meaning of 'Meaning'". In broad strokes, these arguments aim to show that an individual subject's conception surrounding a name does not generally determine what their uses of that name refer to. The primary target is the descriptivist interpretation of Fregean sense, but as we'll see, they also do some work to undermine the generic version of the theory.

Here's Kripke's thought experiment. As Kripke observes, the average speaker is probably incapable of producing a detailed description of Kurt Godel; at best, they can describe him as "the discoverer of the incompleteness theorem." This description does pick out someone uniquely. But we can imagine a situation in which it doesn't pick out *Godel*. Suppose that another unknown mathematician, Schmidt, discovered the incompleteness theorem and Godel stole his work. Even if that were the case, "Godel" would still refer to *Godel*, and not Schmidt. Thus the speaker's conception surrounding "Godel" does not determine the referent of Godel. Instead, Kripke suggests that the referent of "Godel" is fixed by a communal linguistic practice, the details of which need not be privy to individual speakers.

A similar lesson comes from Putnam's Twin Earth. We are to imagine that there's another planet, Twin Earth, that appears to be indistinguishable from earth in all of its macroscopic details, but where the stuff that fills the lakes and oceans, runs through the taps, nourishes life, etc., is made up of a chemical composition distinct from H20. Instead of H20, the stuff that twin earthlings call "water" is made of XYX. The lesson we are supposed to draw from this is that our word "water" refers to H20 and the twin earthling's word "water" refers to XYZ, *despite the fact that the speakers from the two communities will attribute the same properties to the stuff that they each call "water."* (We can imagine this occurs at a time before the 1700s where neither earthlings or twin earthlings have discovered modern chemical theory.) Once again, we gather the lesson that reference is fixed by external factors (e.g. our relations to H20) rather than the individual conceptions that we have of the stuff.

What is the significance of these thought experiments for the Fregean conception of meaning? It shows that *whatever "meaning" a subject grasps, that isn't what determines reference.* Reference is determined independently of a speaker's individual conception surrounding a name. To be specific, it shows that role *1b* cannot coincide with role 3 within a single theory of meaning.

Another lesson can be drawn from Kaplan's discussion of indexicals. Consider the word "I", the first-person pronoun. Understanding this word in English requires understanding a certain rule to the effect that the word "I", when used by a speaker in a context, is to refer to *that speaker*. Kaplan calls this rule the "character" of the expression. It's this rule, the character, that fulfills roles 1a, 1b, 1c and 3 of Fregean sense. However, the character of an expression *cannot* be the feature that fulfills roles 4 and 5. When I say "I am hungry" in a certain context, the proposition expressed is about *me.* The constituent of the thought expressed is *me*. It is not about the "the speaker of the context", which just so happens to be me.

To see this, suppose, for the sake of argument, that the word "I" did contribute its character to the thought expressed by the sentence "I am hungry." Then the sentence, actually uttered by me, would express the proposition <whoever uttered the sentence "I am hungry" is hungry>. Now this is true because I actually spoke the sentence and I am hungry. But suppose, counterfactually, that I wasn't hungry and someone else spoke the sentence. In that case, the sentence "I am hungry" *as actually spoken by me* expressed a proposition *that would have been true,* with respect to that counterfactual circumstance. But that's the wrong result! When I say "I am hungry", the proposition / thought that I express is such that it *will be true in all and only the possible worlds in which Graham is hungry*. The possible worlds where the same sentence is spoken by someone else are irrelevant.

Maybe put it this way. A sentence with an indexical expression, like "I am hungry", has two ways in which it could possibly express a falsehood. In one way, it is uttered by *me* and it is *I* who is possibly not hungry. In the other way, the very same sentence (with the same character) is uttered by someone else who isn't hungry. In the latter case, it would express a falsehood even if *I* (Graham Moore) is hungry. That is because the sentence refers to someone else. In the first way of being possibly false, the sentence expresses a thought about me that could be false. In the second way of being possibly false, the sentence could express a different proposition.

Which proposition a sentence expresses is determined by its character, and the character is not part of the proposition expressed. Thus we must not confuse the character (one type of meaning) with content (the proposition expressed). The problem for Frege, then, is that the concept of sense conflates these.

Frege's principle proves too much

I sometimes wonder whether the classic attacks of Kripke and Putnam could be met by giving a very weak reading of role 3 for senses. The above attacks interpret role 3 as claiming that senses *explain* why a particular referent is fixed. They then proceed to point out that the explanation of reference is independent of what the speaker grasps (Kripke & Putnam) and not part of the proposition / thought expressed (Kaplan, for indexicals).

But what if we drop the idea that senses *explain* reference fixation? We would then have to ditch the descriptivist interpretation of Fregean sense, since that was tailor-made to construe senses as reference fixers. Instead, we could say that words refer to their objects directly*,* unmediated by the help of a Fregean sense, however, propositions are still individuated at a level that's more fine-grained than reference. Indeed, one could say that propositions are individuated at a level that abides by Frege's principle:

(*Frege's Principle)* If a competent speaker who understands sentences *S1* and *S2* can rationally assent to *S1* while dissenting from *S2*, then *S1* and *S2* express different meanings.

We may even call the constituents of such propositions "senses" since they aren't the referents of the pertinent expressions. And if there are any such things, then they must fulfill roles 2 and 4 of Fregean sense. Moreover, we can say that reference *supervenes* on sense in the weak modal sense that doesn't imply explanation. That is, if two words express the same sense, then they are coreferential; but sense does not *explain* how reference gets fixed. This is the weaker reading of role 3.

I gather that this is roughly how Gareth Evans interpreted Fregean sense, as a *de re "way of thinking about the object.*" On this interpretation, the sense of "Hesperus" is *a certain way of thinking of Venus* and the sense of "Phosphorus" is *another certain way of thinking of Venus*. A *way of thinking of X* is ipso facto *about X,* so expressions that share the same sense must share the same referent. But this doesn't imply that the sense itself explains why the referent was determined. If anything, it is one's relation to the referent that explains the sense.

As far as I can tell, this Evansian interpretation of Frege circumvents most (or all?) of the attacks by the externalists. It does, however, come with some cost for the philosophical significance of sense, since it can no longer serve as the theoretical foundation of the view of philosophy as conceptual analysis. But no matter—perhaps it has other attractions. Since the Evansian interpretation upholds Frege's principle, perhaps it offers a more realistic view of the individuation of content in light of speaker behaviour. One could argue that this is important from the point of view of cognitive science.

Alas, even if this interpretation of Fregean sense evades the objections from the externalist, it has another flaw. Unlike the last objection, this flaw is inherent to any theory of meaning that is motivated by Frege's principle. It is thus, to my mind, the most serious objection to Frege's theory of meaning as a whole. The problem is that Frege's principle is in serious tension with any plausible interpretation of role

1 of sense.

Recall, role 1 stipulates that meanings (whatever they are) must explain *synonymy, linguistic understanding,* and *the publicity of meaning.* It's worth mentioning that 'publicity' doesn't only include sharing meanings between two distinct speakers. It also includes sharing meanings between uses of a name by the same speaker over time. My uses of "Hesperus" *now* ought to mean the same thing as my uses of "Hesperus" in the past.

Just as a datum of sheer common sense, it must be true that we often frequently use terms with the same meanings as other people and our past selves. For if we didn't do that, then how would communication be possible? What would language be *for* if not for communicating information?

However, it seems possible *in principle* for *any* two uses of terms that apparently express the same meaning to run afoul with Frege's principle.

Consider any two names "*n1*" and "*n2*" and let's stipulate that they are meant to be synonyms according to the conventional rules of public language. Is it *possible* for some individual speaker of the language to assent to "*n1 = n1*" and dissent from "*n1 = n2*"? Of course! Perhaps they were taught the terms on separate occasions and failed to realize their correference. Perhaps they learned the term "*n1*" through testimony and "*n2*" through direct acquaintance with the referent, and were never given the connection. In that case, they may doubt or dissent from "*n1 = n2*" without thereby displaying any irrationality.

An instance of this story is told by Kripke in his "A puzzle about belief." According to his story, a French man named Pierre learns of the city named "Londres" through the testimony of other francophones. From their testimony, he forms a belief that he would express by "Londres est jolie". On a separate occasion, he goes to London, is told that the city is called "London", and only visits the ugly parts. He thereby forms a belief that he would express as "London is not pretty." If he were to reflect on his beliefs, he would be willing to assent to "London is London" but dissent from "London is Londres". But Pierre has not made any logical mistake; he's not being irrational. This is despite the fact that "London" is the English translation (synonym) of "Londres".

So should we say that "London" and "Londres" aren't *really* correct translations after all, because they differ in Fregean sense (as seen in the cognitive difference exhibited by Pierre)? If we were to do that, we would have to deny synonymy in an awful lot of cases where intuitively we shouldn't. The problem isn't just for words of distinct cultural languages. We could easily concoct a similar case for two distinct words of a common cultural language. (Kripke gives the example of "Paderewski".)

This kind of problem doesn't only affect the terms of public language. A similar point can be made about the names belonging to an individual's language of thought. Suppose that I am gazing at an object, a star in the sky. At one moment I think to myself "that star is bright." The next moment I think to myself "that star is probably dead by now." But then I stop myself: how do I know that the star that I was looking at when I had the first thought is the same as the star that I was looking at when I had the second thought? It may *seem* that I maintained a continuous gaze at the same star throughout the short duration. But it's not (epistemically) *impossible* for me to be wrong. After all, it's possible for one thing to be replaced by another thing without my noticing it. (We can think of radical scenarios where this is possible, perhaps involving evil demons intent on tricking me.) The possibility where I failed to refer to the same thing doesn't even have to be very likely. It just has to be rational in the sense that I'm not betraying any linguistic, logical confusion.[4] It is thus *possible* for me to doubt that "that star" (as used in the first instant)

[4] The Fregean might try to escape the conclusion of this argument by insisting that my doubt *is* irrational after all. In that case, they would be building substantive normative principles into the interpretation of Frege's principle.

shares the same referent with “that star” (as used in the second instant), without me irrationally contradicting myself. According to Frege’s principle, that’s sufficient to conclude that the two terms have different senses. Well, if that meagre possibility of me doubting (without irrational self-contradiction) that “that star [first use] = that star [second use]” is enough to prove that they have different senses, then distinct uses of words will hardly ever have the same sense.

This consideration shows that there is a deep conflict between two desiderata for Fregean sense. On the one hand, we want Fregean senses to explain the intersubjectivity of meaning. They are supposed to explain how two subjects (or two timeslices of the same person) can mean the same thing with their words, even if they do not have exactly the same cognitive relation to the referent. This pushes us towards a course-grained conception of content. On the other hand, Fregean senses are supposed to record the cognitive situation of the subject vis-a-vis the referent. If taken as sacrosanct, this pushes us towards a highly fine-grained conception of content, to the point where synonymy is practically unattainable.

At this point, most of us reasonable folk will conclude that Frege’s principle is false. Whatever meaning is, it should not be so construed as being so fine-grained that it tracks all possible (self-consistent, rational) assents and dissents.

If we admit that much, then that spells serious trouble for the *motivation* for Frege’s conception of meaning. Remember, the argument for the existence of Fregean sense stemmed from a natural reaction to Frege’s puzzle concerning “Hesperus” and “Phosphorus.” But to convert the intuitive response into a conclusion about *semantics* and *meaning*, we needed to appeal to Frege’s principle. Without that principle we no longer have any reason to include cognitive significance (role 2) as an element of a theory of meaning. And without that role, Frege’s theory loses a significant amount of its appeal.

There was in fact one philosopher who upheld something like Frege’s principle and took it to its natural conclusion. I’m talking about the descriptivism of early Bertrand Russell. Like Frege, Russell also thought that a speaker ought to have infallible access to the facts of meaning for their own speech. Specifically, they must know, with cartesian certainty, whether any two sentences possess the same meaning or not. From these principles, Russell was notoriously driven to the conclusion that the meanings of a subject’s expressions are highly individual to them and fleeting in time. The only real proper names in Russell’s scheme are ones that refer to the speaker’s private sense data in the present. Most commentators have regarded Russell’s picture of meaning as a disaster, since its extreme individualism is far removed from anything resembling a language. Nonetheless, this is the natural consequence of constraining meaning by Frege’s principle.

We can summarize the argument this way. Frege’s theory of sense is primarily motivated by Frege’s principle. But it follows from Frege’s principle that senses must be so fine-grained that they couldn’t be public, thus forfeiting their claim to be a kind of meaning. Hence Frege’s principle is false, and the motivation for Frege’s theory loses its force.

Originally, the import of “rational” in Frege’s principle is supposed to mean something like “without logical error.” But if we give a souped-up reading of the principle, one might insist that, for broad epistemic reasons that need not be transparent to the subject, it may be rational to dissent from “Hesperus is Phosphorus” but it is always irrational to doubt that “that star [in one moment] is that star [in the next]”. Likewise, it is always irrational to dissent from “London is Londres”, since they’re synonyms, *even when the correferentiality is unknown to the subject.* This view will then deliver the result that “Hesperus” and “Phosphorus” differ in sense, but the other pairs of examples do not.

I’m vaguely aware of some examples of theorists who take this route whom I won’t name because I’m not sure. But the end result of this is that *meaning,* on the layer of Fregean sense, gets carved up according to the principles of substantive rationality. It’s an interesting view. I just don’t buy it because I don’t think that Pierre is being irrational.

Conclusion

Where do these considerations leave us? This doesn't yet go all of the way to proving that the theory of direct reference is true. After all, Fregeanism and the theory of direct reference are high-level philosophical approaches to semantics. It's doubtful that any one argument will be decisive. The best we can do is amass the considerations for and against each one, and see which carries more weight. (Not only that, but as I have mentioned a few times, the list of roles fit for a theory of meaning is somewhat up for grabs.)

However, the previous consideration does show that the apparently compelling motivation for Fregeanism, based on the Frege puzzles, is illusory. Not all differences in cognitive significance track differences in (publicly available) semantic content. Therefore, the difference between "Hesperus" and "Phosphorus" need not count as a semantic difference.

A defender of Frege could, at this point, *insist* that the "Hesperus" and "Phosphorus" do differ in meaning, even though the other cases (e.g. "London" and "Londres") don't. But it's hard to see how a principled line between them could be drawn if it's not based on Frege's principle.[5] Ascribing a difference in sense between some, but not all, differences in cognitive significance is a very uneasy place for a Fregean position to rest.

If this is the state of the dialectic, then there is one more consideration that tips the balance in favour of the theory of direct reference. Basically, the theory of direct reference has another advantage that the (non-descriptivist) versions of Fregeanism lack. Namely, the theory of direct reference offers a straightforward explanation of another prized desideratum for a theory of meaning. The theory of direct reference can explain how meaning is *compositional*.

It has been mentioned a few times that there is a theoretical constraint on any theory of meaning that meaning is compositional. Basically, the meanings of words ought to combine, in some systematic fashion, to generate the meanings of sentences. This fact about meaning is important for explaining how any speaker can understand a potential infinity of sentences by grasping only finitely many words. Since it is an empirical fact about us that we have this ability, this represents a fairly non-negotiable feature of any viable theory of meaning.

With that said, here is an argument for direct reference theory. First, it is possible to explain the compositionality of meaning when meaning is understood as reference. (There is a robust literature of formal semantic theories that identifies sentence meaning with truth conditions and then explains their composition on the basis of the 'denotations' of lexical items—including reference for names.) On the flip side, it is relatively obscure how meaning could be compositional if the meaning of names is more fine-grained than reference. Therefore, if all else is equal, we ought to identify the meanings of names with their referents.[6]

[5] Or, as per footnote 4, they could insist that it is *rational* to doubt that "Hesperus is Phosphorus" but *irrational* to doubt that "London is Londres" according to some substantive principles of rationality that are settled pre-semantically. It's just very difficult to see how such a view could be anything other than ad hoc.

[6] This argument is actually qutie a bit less straightforward than I have led on. We have known for centuries that belief contexts are a powerful source of counterexamples to reference-based compositional semantics. Almost everyone agrees that "S believes that Hesperus is Hesperus" can be true while "S believes that Hesperus is Phosphorus" is false. The main challenge for the direct reference theorist is to deal with this alleged fact. (Indeed, the compositionality of meaning in belief contexts is one of Frege's initial motivations for his view!) And while

The previous section argues that pretty much all else is equal. Thus the conclusion leaves things with the weight of evidence favouring the theory of direct reference.

we're at it, we should also explain how we can coherently describe the belief states of a subject who is in the grips of a Frege puzzle. Salmon, in his book *Frege's Puzzle,* goes a significant way to dealing with these problems.