Fregeanism and Mental Content

Graham Seth Moore
December 2023

As I understand it, Frege's two-tiered theory of meaning introduces both sense and reference as kinds of *theoretical entities*—that is, they are each posited as fulfillers of their own respective theoretic roles. Reference is supposed to play the role of explaining semantic composition via functions and objects. This is why, for Frege, *sentences* have reference (their truth values). Treating truth values as referents allows *reference* to play its role in explaining the semantic composition of sentences by *truth* functions defined over *referents*.

Sense, in this picture, is also defined primarily through its theoretical roles. In "Sense and Reference", I discern at least five distinct roles for sense:

> (1) senses are a kind of meaning of linguistic expressions (they explain synonymy, understanding, publicity, communication);
>
> (2) Senses are supposed to explain sameness and difference of *cognitive significance*: (e.g. why "Hesperus is Hesperus" is trivial while "Hesperus is Phosphorus" is informative);
>
> (3) Senses determine reference. (This has a weak version: reference supervenes on sense — terms with different referents must have different senses. It also has a strong version: sense encodes the mechanism by which reference is fixed — for example, the descriptions that determine reference through satisfaction.)
>
> (4) Senses are the constituents of thought. This explains why thinking that *Hesperus is a planet* appears to be distinct from thinking that *Phosphorus* is a planet; these thoughts *really are* distinct because they are composed of different objects.
>
> (5) Senses serve as the referents of expressions within propositional attitude contexts (Frege's theory of indirect reference).

I take it that these roles are essential to what I'll call "classical Fregeanism". According to classical Fregeanism, there's a perfect coincidence between linguistic and mental content, since one and the same thing is supposed to play both roles (1) and (4).[1] This means that language and propositional attitudes are theorized together. What a subject *says* by an utterance is identical to what is entertained in thought.

Role (2) entails the distinctive Fregean doctrine that content is more fine-grained than reference. Hence, "Hesperus" makes a different contribution to content than does "Phosphorus", despite the two sharing the same referent.

[1] Interestingly, Frege makes an exception for a thinker's own idea of themself. He writes that the sense of my mental indexical <I> is something that only I can grasp. Hence, it is not communicable through public language. When I tell you that "I am tired", the sense expressed by the public linguistic term "I" is some mutually graspable description, like "the person speaking"; it is not the sense of my mental indexical.

I think it's fair to say that (2) has both a descriptive, psychological interpretation and a normative, epistemological interpretation. According to the descriptive, psychological interpretation, to say that <Hesperus is a planet> differs in *cognitive significance* from <Phosphorus is a planet> is just to say that the subject's cognitive system is tokening naturally distinct state types when one or the other is thought. Perhaps this is because the subject is exercising distinct cognitive capacities to distinguish objects (as in Evans 1982). Here, *sense* is treated as the natural way to carve cognitive reality at its joints; it is not (at first) to impose our pretheoretical judgments about what's *rational.*

On the alternative, normative epistemological interpretation, *sense* is tailored specifically to explain rationality. Here, it's taken as a starting point that it's possible to rationally doubt that *Hesperus is Phosphorus* without doubting that *Hesperus is Hesperus*. The Fregean then posits a level of meaning that lines up with these judgments of rationality. Under this conception, content is individuated by a normative version of Evans' intuitive principle of difference:

> (*Frege's Principle*) <P> and <Q> differ in content if and only if it's possible for a rational subject to take conflicting attitudes towards <P> and <Q> (believing <P> while doubting <Q>).

Indeed, something like this principle is assumed in the usual arguments for Fregeanism from the Frege cases.

1. Problems with classical Fregeanism.

I've already written a long post on the arguments against classical Fregeanism. In retrospect, a number of the classics (e.g. Putnam and Kripke's externalist arguments against descriptivism, Kaplan's treatment of context-sensitivity) really only show that the *strong reading* of (4) doesn't cohere with (1). That's just to say that the mechanisms that explain reference (e.g. causal chains, Kaplanian character) should not be conflated with *content.* Or, to put it in more contemporary terminology, there is a distinction between *semantics* (what a word contributed to meaning) and *metasemantics* (what explains how a word is endowed with its meaning or referent). The strong reading of (4) threatens to conflate them.

But there's always the weak reading of (4) to fall back on. On this view, the Fregean only claims that words with different referents always have different senses. There's no claim that the sense of an expression plays the metasemantic role of explaining reference fixation. (If one takes this weaker line, one would probably have to give up classic descriptivism and go in for *de re* senses.)

Nonetheless, even if those arguments are evadable, there is still the much more substantive challenge posed by Kripke's *Paderewski*. According to the story, Peter hears the *public language* name "Paderewski" from two different sources; they respectively say "Paderewski is a politician" and "Paderewski is a musician." Well, *if*, as per role 1, sense is supposed to explain *meaning in public languages*, then "Paderewski" should have the same sense in each occasion. However, it's intuitively possible for Peter to fail to realize that the two uses of "Paderewski" were referring to the same guy. If so, then it is possible for Peter to *doubt* that *Paderewski the politician is Paderewski the musician.* If, as per role 2, sense is supposed to explain cognitive significance, then the two uses of "Paderewski" should have different senses. Moreover, if it isn't *irrational* for Peter to have this doubt (alternatively, if it would be *irrational* for him to infer that someone is both a politician and musician), then the two uses of "Paderewski" should have different senses on the normative epistemological interpretation of 'sense.'

To summarize, whatever explains *meaning in a public language* does not coincide with what explains *individual cognitive significance and rationality.* Roles (1) and (2) cannot be jointly satisfied by the same theoretical entity.

2. Three versions of Frege's puzzle

Here's a reaction I've heard before. One might say "okay, fine, maybe the semantics of *public language expressions* isn't Fregean. Be that as it may. But what I really care about is *rationality* (or more generally, how semantics affects epistemology). And rationality is a matter of the *attitudes* of a subject: their beliefs, desires, thoughts, and intentions. It's *these* for which the Fregean sense/reference distinction is most important, not the common currency languages like English, French, etc."

In what follows, I would like to focus on a variant of Fregeanism that takes this line. That is, it will apply the sense/reference distinction to the *attitudes* only, and not to public language. (There is precedent. I believe Evans is a prime example if I remember him correctly.) The basic idea would be to hold that mental *contents* are individuated by sense, not reference (role 4) and this means that contents track sameness and difference in cognitive significance (role 2). Such a view will depart from classical Fregeanism by denying that roles (1) and (4) coincide; it allows mental content to be more fine-grained than linguistic meaning. So it isn't really Frege to the letter. However, there is still something distinctly Fregean about it. It still holds that Frege was basically right in his reaction to something that's basically like Frege's puzzle.

In order to clarify the last claim, we must also be clear about Frege puzzles. I think it's fair to say that there are *at least three* distinct puzzles that arise from Hesperus/Phosphorus cases.

Puzzle 1. The first kind of puzzle is the one that Frege actually presented. The opening paragraphs of "On Sense and Reference" specifically address *sentences*. "Hesperus is Hesperus" and "Hesperus is Phosphorus" are distinct *sentences* with apparent differences in the informational content they semantically encoded. The first puzzle is to explain this difference in felt cognitive significance.

The classic Fregean solution is to say that there *appears* to be a difference in their informational content *because there is* a difference in their informational content. This solution is first and foremost about the entities that fulfill role (1), and so anything that follows will tie role (1) to sense. The classic Millian solution is to say that the apparent difference in informational content is actually a matter of pragmatics, not semantics (Salmon 1986). This, however, is only a claim about the information of public language expressions; it is consistent with virtually any claim about the attitudes.

Puzzle 2. The second puzzle concerns the semantics of attitudinal ascriptions. Specifically, our semantic theory of attitude ascriptions must explain our dispositions to assent and dissent with regards to reports of the form "S believes that Hesperus is F" and "S believes that Phosphorus is F". Speakers are often willing to assent to one and dissent from the other. If so, then what do these sentences mean?

There are really two problems here. The first is motivated by the observation that ordinary speakers do not mean to attribute flagrant irrationality to a subject when they (the speakers) simultaneously assent to "S believes that Hesperus is F" and "S believes that Phosphorus is not F". The second is motivated by the observation that ordinary speakers are not themselves flagrantly irrational (contradicting *themselves*) when they assent to "S believes that Hesperus is F" and dissent from "S believes that Phosphorus is F".

Notice, though, that this puzzle is still about the semantics of public language. It just concerns the narrow part of public language where statements are embedded after the verb "believes".

The classic Fregean solution is his theory of indirect reference. After the verb "believes", words cease to refer to their ordinary referent, and instead refer to their customary sense.

There are a few Millian solutions in the literature, but the one I'll mention here is Salmon (1986)'s "guise Millianism". The basic tactic is to treat the verb "believes" as semantically equivalent to a three-place predicate. It takes as values a subject, a Russellian proposition, and a "guise". A guise is kind of like a mode of presentation. It's like how the proposition presents itself to a subject. So according to the guise Millian, the real logical form of "S believes that P" is "S believes x under guise m". This allows "S believes that Hesperus is a planet" and "S believes that Phosphorus is a planet" to have the same informational content in the 'x' slot, but express different propositions by dint of taking different values for 'm'.

Puzzle 3. For the third puzzle, we finally forget about language and focus on the mental states themselves. Take a subject who harbours beliefs (/ thoughts) about Hesperus / Phosphorus, but fails to realize that the planet they see in the morning (in one 584 day period) is the same as the planet they see in the evening (in the next 584 day period). Under these circumstances, intuitively, we would not want to count their <Hesperus>-thoughts as identical to their <Phosphorus>-thoughts. One wants to say that they may believe <Hesperus is a planet> without believing <Phosphorus is a planet>. Perhaps this is owing to rationality considerations, or perhaps it's motivated by descriptive psychology. Either way, there's pressure to count these thoughts as separate in one's metaphysical picture of cognition.

3. Guise Millianism, Fodor's RTM

My concern is puzzle 3. But first let me say how I got here. I was reading François Recanati's *Mental Files* and I came across a passage where he claims that Fregeanism about mental content and guise Millianism are different in name only—that the dispute between them is merely verbal. He doesn't mention Nathan Salmon by name, but he writes of a position that Russellian propositions (propositions composed of worldly objects) are "believed or disbelieved only under guises" (8). He then writes that this option

> "amounts to a concession of defeat; for guises are nothing but modes of presentation, and modes of presentation are now allowed to enter into finer-grained propositions construed as the *complete* content of the attitudes. Far from conflicting with Frege's construal of propositions as involving senses, this view merely introduces a new, coarser-grained notion of 'proposition', namely R-propositions, playing a different role and corresponding roughly to an equivalence class of Fregean propositions. This is a variant of Frege's two-level approach rather than a genuine alternative of the sort Russell was after" (8).

This passage awoke me from some kind of dogmatic slumber. I *thought* I understood what was at stake between Fregean and Russellianism when transposed away from language to the cognitive environment. But retracing my steps convinced me that the matter was more complicated than I thought. So that's the problem that I want to raise and answer in the rest of this paper.

There is some pedantry that must be cleared up. *Russellianism* is a position about the individuation of *propositions*. It says that propositions are individuated coarsely by worldly objects. Hence <Hesperus is Hesperus> and <Hesperus is Phosphorus> are identical propositions. *Millianism* is a position on *names*. It says that the meaning of a name is its referent. If Millianism is true, then the meaning of "Hesperus is Phosphorus" is a Russellian proposition (given uncontroversial assumptions about compositionality). But still, we must continue to distinguish claims about language and claims about attitudinal content (propositions).

I say this because it would be a mistake to take Recanati's comment to pertain directly to the "guise Millianism" of Salmon (1986). (It's thus a good thing Recanati doesn't mention Salmon.) Guise Millianism is, first and foremost, a *semantic* theory. It says that belief *ascriptions* are disguised three-place predicates. It is only *directly* a solution to puzzle 2, not a direct solution to puzzle 3. We must distinguish between semantic claims about belief reports and the metaphysics of belief. After all, there's no rule that says that natural language expressions must semantically encode the same adicy as the qualities to which they correspond. (Indeed, there's some reason to expect mismatches. See Kenneth Taylor (2019) chapter 6.) So it's quite consistent to claim that the correct semantic representation of "belief" is that it's a three-place predicate, without also claiming that belief itself is a three-place relation.

(The same is true of Frege's theory of indirect reference. It is primarily a semantic theory, not a theory about belief per se, although Frege's followers often run them together. This is the linguistic turn at work.)

Now that that's out of the way, let's put guise Millianism aside and consider a position that *actually does* claim that attitudes are three-place relations. This view, of course, is Jerry Fodor's *representational theory of mind*. Indeed, sometime around the late 80s early 90s (especially in *The Elm and the Expert*), Fodor took the coarse-grained view of content that I wish to contrast with Frege's.

Here are the essentials of the view. As advertised, Fodor takes propositional attitudes to be three place relations between a subject, a representation, and its content. Thus the fact corresponding to a belief report, *"S believes that P"*, is the fact that the subject S bears a (functionally characterized) attitudinal relation to a token of their internal representation *m,* and *m* semantically encodes the (Russellian) proposition *that p*.

I'm going to say more about this shortly. But before I do, I first need to explain why I'm worried. As I understand Fodor's theory, reference will supervene on mental representations. That is to say, there's no difference in reference without a difference in representation. However, there may be different representations that have the same reference. Thus representations are more fine-grained than reference.

As everyone knows, Frege has a two-level semantic theory. There's sense and there's reference. The subject bears the belief relation to a thought. The thought is individuated by fine-grained sense. In the Fregean paradigm, we call the thought the "content" of the belief. The "content" has the role of explaining, *inter alia,* cognitive significance. The thought also *has* reference (it refers to a truth value), as do its parts. Reference plays, among other things, the role of explaining semantic composition.

Fodor also has a two-level theory. There's mental representations and there's content. The subject bears a belief-like relation to a representation. The representation is fine grained. In the Fodorian paradigm, we call the representation the "vehicle" of content. It has the role of explaining, *inter alia*, cognitive *processes.* The vehicle/representation also *has* content, as do its parts. Content plays, among other things, the role of explaining semantic composition. Finally, content, according to post-1990 Fodor, is individuated at the level of reference.

I hope you see what I'm getting at here. There are differences, to be sure; Fodorian mental

representations play more specific roles than I've described so far. They are also supposed to be "syntactic" rather than semantic. But even so, it's consistent with the position to individuate representations by cognitive significance. Fodor even calls the representations "modes of presentation" (in *The Elm and the Expert*). In that case, the <Hesperus> representation would be distinct from the <Phosphorus> representation within the cognitive system of a subject who isn't privy to their coreference. Well, in that case, the last two paragraphs should start to look isomorphic. The only surface-level difference is that the Fregean and the Fodorion use the word "content" for different things. The Fregean uses "content" to label the thing that plays the fine-grained role of explaining cognitive significance; and the Fodorion uses "content" to label the thing that plays the coarse-grained role of explaining semantic composition. But each of them can agree that there's *some fine-grained thing* that plays the role of explaining cognitive significance and some coarse-grained thing that plays the role of explaining semantic composition. Which of these to call "content" is not a substantive issue.[2] Is there any real distance between them?

I think there is. But before looking under the hood, I want to set aside one totally uninteresting answer. Again, as everyone knows, the original Frege thought that senses were platonic abstract entities, residing in an otherworldly realm, not in the mind of any thinker. Fodorian representations, by contrast, are located somewhere in our brains. That's a significant difference … right?

Wrong. It's true that Frege thought that. But the Fregean tradition survived the subsequent identification of senses with concepts in the minds of thinkers. The real issue isn't so much the *entities* that compose content, or where they are located. (Talk of propositions being 'composed of' worldly entities in the Russellian tradition was always a little spooky.) The real issue is attitude individuation. Suppose that Harry believes that Hesperus is a planet and Gary believes that Phosphorus is a planet. The real question is: do their beliefs count as the same or not?

4. The roles of content

There is no answering this question without revisiting the notion of 'content.' What is content anyway? What do we need it for?

To echo the opening sentences of this essay, I have no grip on this notion unless by understanding it as a theoretical one. By this, I mean that we define 'content' by listing the roles that content is supposed to play, and taking it to be whatever entity satisfies those roles. This practice is overtly stipulative, but it's surrounded by substantive issues, which people can meaningfully disagree over. Once we have the list of theoretical roles, we can ask such questions as: can we expect one entity to play all of them? Do they cohere? Do they form a natural cluster?

What are the theoretical roles that define our notion of content? At least for beliefs about ordinary objects, one role reigns supreme. The reason we take people (organisms, cognitive systems) to have attitudes with mental content is because it explains their *actions* and *behaviour*. Specifically, it explains a person's successful and unsuccessful navigation with distal objects. (For a nice articulation of this role, see the third chapter of Shea (2018) and the first chapter of Fodor's *Psychosemantics* (1987).)

Imagine that you blindfold two people and have them run through a field to retrieve a prize. One of them runs 50 steps, turns left, and finds the prize. The other one runs 50 steps, turns right, and falls into

[2] Recanati considers Fodor's position (and Tye and Sainsbury's) on page 245 of *Mental Files* (2012) and explicitly claims that it's a 'terminological variant.' Although he then gives some reason for thinking that the vehicles are more than merely syntactic objects, which I'll consider in an appendix.

a pit. Why did the first person succeed and the second one fail? Among other things, we say that the first person *believed* that the prize was to be found by turning left after 50 steps—which was true—and the second one *believed* that the prize was to be found by turning right after 50 steps—which was false. Different beliefs lead to different outcomes, and different truth values lead to a difference in success.

This first role immediately implies a second. For beliefs to play this role in action explanation, they must be capable of truth and falsity. Content explains this because according to this theory of mind, a belief is true (/ false) if, and only if, its content is true (/ false). Hence content, whatever it is, is a truth bearer.

Bearing truth/falsity means that contents are semantic entities. Being semantic in nature means that contents must be susceptible to compositionality constraints. For example, if a subject believes that *P & Q*, then their belief is *true* if and only if the content *that P* is true and the content *that Q* is true. Analogous principles must hold for the other operations of combining whole contents, along with the principles for forming atomic truth-evaluable contents from sub-propositional ideas and concepts.

(In *Concepts* (1998) and *LOT2* (2008), Fodor took the compositionality principle to provide a positive argument in favour of Russellian content over Fregean content. The basic idea is that we have some handle on how reference composes but no handle on how senses compose. Sure, Frege *claimed* that senses compose. And sure, the *descriptivist* interpretation of sense will provide an explanation of how senses compose. But if we reject descriptivism in favour of seeing senses as stereotypes, or clusters of information, or inferential roles, or whatever, then it's hard to uphold the compositionality constraint on content.)

Does content have any other roles? Sure. It has some role in fixing the meanings of a subject's linguistic utterances. When I utter "that is a very tall mountain" while looking out my window, the referent of my demonstrative is fixed in part by the *content* of my perceptual state (and perhaps something about *attention* and the *content* of my intentions for how to be interpreted by my audience).

Content also has a role in explaining agreement and disagreement between subjects. We reach *genuine* agreement when the contents of our beliefs (in question) is the same; we have a genuine *disagreement* if the content of my belief is the content of your disbelief. Our agreement or disagreement is *apparent* if there is no such match in content.

I think that this much is all common terrain. What remains is the role that I think is contested. Specifically, we haven't yet given voice to the theoretical ambitions of the kind of Fregean I quoted earlier. Remember what they said;

> "what I really care about is *rationality* (or more generally, how semantics affects epistemology). And rationality is a matter of the *attitudes* of a subject: their beliefs, desires, thoughts, and intentions. It's *these* for which the Fregean sense/reference distinction is most important,"

There's a kind of thinker—indeed, *many* philosophers—who expect our metaphysics of mind to carve at the epistemological joints. We might say that for them, content has the *additional* role of explaining rationality and irrationality. It does this by content individuation. Roughly, a subject is *irrational* for believing <*P*> and believing <~*Q*> *if* content <P> is identical to content <*Q*>. However, a subject who believes that <Hesperus is a planet> and that <Phosphorus is not a planet> *may not* be irrational if <Hesperus is planet> and <Phosphorus is a planet> are distinct.

I have no doubt that this theoretical role for content is what animates a great deal of Fregeans. For instance, it's written right into the Fregeanism of Jarvis & Ichikawa (2016), which takes senses to be

identical to inferential roles, and inferential roles to be defined by how a subject *ought* to infer (not by how they *actually* infer). However, I noted earlier that this role may not be definitive of all Fregeans. (I seem to recall that Evans is more interested in descriptive psychological explanation. And as I'll explain in the appendix, Recanati is more interested in the phenomena of indexicality.)

From now on, I'll take this idea as fairly central to Fregeanism (even if it's not universal). The question now is whether mental representations, in the RTM, are up to the task.

5. The role of vehicles

Content plays the role of explaining a subject's successful/unsuccessful actions in their navigation with distal objects. According to the representational theory of mind, *mental representations (vehicles of content)* play the role of *locally mediating* the cognitive processes required for such navigation.

Suppose I have the representation <Momo is on the window sill or Momo is under the bed> deployable in my cognitive system in the manner distinctive of belief (or to put it in lay terms: I believe it). Through perception, a new representation enters the system, <It's not the case that Momo is on the window sill>. From these, my internal processes get to work. They've been programmed to take input representations that look like <A or B> and <Not A> and then output the representation <B>. Hence my cognitive system produces the new representation <Momo is under the bed> and deploys it in the way that's distinctive of belief. I march over to the bed, look underneath, and find Momo there.

According to the representational theory of mind, all of this is possible because I have internal representations, <Momo>, <on the window sill>, <under the bed> that track Momo, things on the window sill, and things under the bed, respectively. Moreover, I have internal processes that are programmed to manipulate these representations based on their syntactic features. However, even though these internal processes only "see" the internal features of these mental representations (their syntax), they faithfully preserve semantic properties that relate them to the things represented. We can design a machine that outputs <B> from any inputs <A or B> and <Not A>. The machine may only "see" the syntax, but its manipulations will preserve truth from inputs to output. (This can be proven using truth tables.)

This, then, is the role of mental representations in the representational theory of mind. They are the internal vehicles that *carry* semantic information *through* cognitive processes that are programmed to manipulate syntax.

6. The central question

I've now given enough description to raise the question that I think separates the typical Fregean from the typical Fodorian. We started with the question: "Can mental representations play the role of Fregean sense?". If the answer is yes, then Recanati is right and the RTM is just a "terminological variant" of Fregeanism embedded in a computational understanding of cognition. If the answer is no, then the combination of the RTM with a referential semantics really is more in line with the Russellian tradition.

I stipulate that the "role of Fregean sense" of interest here is explaining rationality. Mental representations are the internal medium of cognitive processes. So now we raise the substantive question: Are the internal vehicles that underwrite our cognitive processes individuated into types that explain our intuitive judgments of rationality/irrationality (esp. in Frege cases)?

In a way, this really is just the age-old question of whether our metaphysics of mind is beholden to epistemology. When we inquire into the nature of the mind, should we presuppose from the start that its

nature (particularly, in how we individuate mental vehicles) will uphold our practice of judging others rational and irrational?

7. Some initial reason to be hopeful

Actually, there's some reason to think, at first, that whatever plays the syntactic vehicle role is needed to explain some aspects of rationality. (I've often heard this argument attributed to John Campbell). Suppose an agent performs the following inference in thought:

(1) *a* is F
(2) *a* is G
(C) Something is both F and G.

Intuitively, drawing this inference is a paradigmatic example of rational behavior. But there's a catch. It *presupposes* that the token of <a> in the first premise has the same referent as the token of <a> in the second premise. If we mark them off as distinct tokens, so as not to prejudge this presupposition, then the rationality of this inference is no longer apparent:

(1) a_1 is F
(2) a_2 is G
(C) Something is both F and G.

This makes clear that the inference goes through *only if* we assume that a_1 is identical to a_2.

Suppose that we tried to incorporate this presupposition as an added premise to our inference. In that case, the real piece of reasoning would rather be:

(1) a_1 is F
(2) a_2 is G
(3) $a_1 = a_2$
(C) Something is both F and G.

However, this just pushes the question back. This new inference only goes through if we can presuppose that the token of $<a_1>$ in premise 1 correfers with the token of $<a_1>$ in premise 3, and if the token of $<a_2>$ in premise 2 corefers with the token of $<a_2>$ in premise 3. Should we incorporate these two presuppositions as two more added premises? That way clearly leads to an infinite regress, and the rationality of the initial inference would never get explained.

To explain rationality, there must be some condition on the pair of tokens of <a> (or $<a_1>$ and $<a_2>$) that *coordinate* them in such a way that permits the inference above, and this condition need not be represented as an explicit premise within the subject's reasoning.

Should we say that the inference is rational if $<a_1>$ and $<a_2>$ are *in fact* coreferential (coreferential *de facto*)? I myself tend to have a higher tolerance for epistemic externalism than the average person, but even I'll admit that there's a mountain of intuition against this. Take a subject who has no reason to believe that the heavenly body he sees in the evening is the same as the heavenly body he sees in the morning. And suppose he spuriously inferred:

(1) Hesperus is a rocky planet
(2) Phosphorus is the second planet from the sun
(C) There is a rocky planet that is second from the sun.

Given that our subject has no clue that the planet he represents as <Hesperus> is the same planet as he represents as <Phosphorus>, this inference is patently irrational, *despite* the *de facto* coreference of <Hesperus> and <Phosphorus>.[3] D*e facto* correference cannot be the right condition for token coordination for rational inference.

If this is right, then the condition that makes this inference rational cannot be found at the level of reference. Rather, it must pertain to the token representations that appear in each premise. Let's say that two token representations are *de jure* coreferential if they *purport* to be coreferential according to the set up of the subject's representational system. By this definition, an inference like the above is rational if the tokens of <a> are *de jure* coreferential. The remaining question is what constitutes *de jure* coreferentiality.

The representational theory of mind gives us a straightforward answer (or so it seems): two tokens are *de jure* coreferential if, and only if, they are tokens of the same representation *type*. What does it mean to say that two tokens are of the same representation type? According to early Fodor, it's because they share the same *syntax.*

According to this idea, each name in the language of thought has a unique syntactic representation. Multiple instantiation of syntactic type is how the representation system encodes *de jure* coreferentiality. Finally, *de jure* coreferentiality explains rational inference. The upshot is that inferences in the language of thought can be classified by the same syntactic criteria that determine the validity of arguments in the symbolic languages of formal logic. For instance,

(1) *a* is F
(2) *a* is G
(C) Something is both F and G

is rational, whereas

(1) *a* is F
(2) *b* is G
(C) Something is both F and G.

is not (not without an added premise that $a = b$).

Since the individuation of names in the language of thought plays this role of explaining *de jure* coreferentiality for rational coordination, there's something very Fregean about this whole idea. We might say that names are individuated in the language of thought according to *sense*, not reference. Indeed, Aidan Gray (forthcoming) remarks how formal symbolic languages are designed to model Frege's sense/reference distinction by representing sense by syntactic sameness.

Result: the score is now 1-0 for the Fregeans.

But alas, the idea that rational coordination is encoded by syntactic sameness (which corresponds

[3] I don't mean to suggest that the failure of rational inference is due to the subject's lack of metarepresentational knowledge about what his tokens refer to. I'm just trying to neutrally convey that the subject is in a Frege case.

to sense) faces a number of problems. For one, Fodor himself abandoned the idea that representation types in the language of thought are individuated by syntactic properties. (This idea leaves little hope that different subjects can share the same concepts, since there's no way that my concept of Hesperus is realized by the same neuro-physical type as your concept of Hesperus.) Instead, Fodor (e.g. in *LOT2* (2008)) spoke of concept types being individuated "functionally", which is to say that we don't really know what we're talking about. I wrote about this issue here: *A quick argument for the originalist theory of concepts*. There I argued that concept individuation does not supervene on features that are *intrinsic* to the tokens (like syntax). I'm not suggesting that this is a knock-down objection to the idea that the representations in the RTM encode Fregean senses. But to explore the issue any further would get exponentially more complicated—more complicated than I care to explore right now.

8. Some reason to be doubtful

Besides, there's another problem for the Fregean that may be more serious. It's a contemporary variant of Kripke's "Paderewski". It cuts through the idea that rational coordination between tokens can be explained by sameness of *anything*.

First let me say a word about concepts. Concepts, according to the Fodorian paradigm, are representational *vehicles*. You might think of them as the "words" in the language of thought. In this picture, concepts *have* content; indeed, according to post-1990 Fodor, the content of a concept is its reference. Now Frege himself reserved the word "concept" for something technical: a function that outputs truth values. (First-level concepts are functions from objects to truth values; second-level concepts are functions from first-level concepts to truth values; and so on.) But let's put that technical use of the term "concept" aside. Really, in the wider Fregean tradition, concepts are broadly used to refer to the constituents of thoughts, and thoughts *are* content. Some Fregeans reserve the term "concept" just for the predicative kind; others use the term for all syntactic categories. I'm going to use the word in the simplest way and call all kinds of thought constituents "concepts." In particular, I'll use "concept" to refer to the constituents of thoughts that are the mental analogue of names.

The Fregean position may thus be expressed as the claim that concept individuation undergirds rational coordination.

However, it turns out that this is in tension with another plausible thesis about concepts. It's called *anti-individualism*. Here, I understand anti-individualism as the thesis that concepts are not always endogenous to an individual. Concepts can be *transferred through communication*.

Consider any close friend you have that really likes to gossip. Since he's a close friend, he's exhausted all of his avenues of gossip about people you're mutually acquainted with. But since he can't help himself, he continues to gossip about people whom you've never heard of before. Imagine he goes on and on about John, that asshole from the office whom he suspects is cheating on his partner.

I take it that this sort of thing makes it possible for you, the audience to your friend's gossip, to become endowed with a concept of John. By this I mean that you can have thoughts *about John*—in addition to the merely descriptive general thoughts concerning "the asshole named 'John' that my friend was referring to—whoever that is."

Whatever concepts are, they are the constituents of thoughts. They originate in the minds of a subject when that subject produces their first token, often in response to a novel experience. From then on, the subject continues to produce tokens whose type is determined by *deference* to past tokens. So, for instance, my token <Socrates> that I instantiate right now is a token of the concept type <Socrates>

(rather than an original use of a new concept) because my cognitive system conjured it (or its vehicle) up from the <Socrates> file, which has a certain history. The present token's identity is determined by that history.

I don't really know how to talk about this 'deference' relation (that's a problem for thinkers more serious than I), but regardless, the main suggestion of the previous two paragraphs is that this deference relation can jump subjects. When I think of John, I'm not *creating a new concept* (through coinage via the description "the asshole named 'John' that my friend was referring to"); rather, I'm tokening an *old concept* that originated in a different subject. The identity of my concept is determined through deference to my friend. Or, to put the same point in a way that's less truistic than it first seems: language is the medium through which we share concepts.

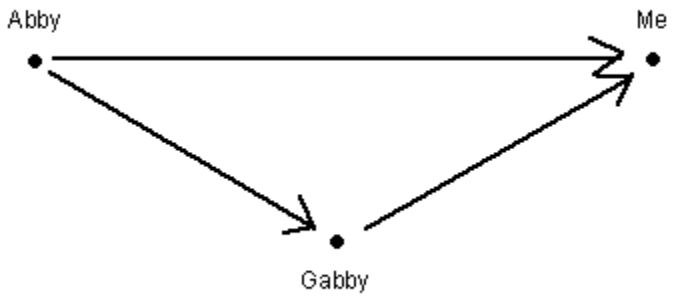
Once we accept this anti-individualism thesis, we can construct the following Paderewski case. (Note: I'm varying it slightly in a way that makes most sense to me.)

Abby is a gossip who is also the original acquaintance of Paderewski. Abby gossips to Gabby, who doesn't know Paderewski personally, but thanks to the anti-individualism thesis, is bestowed with a concept of Paderewski that defers back to Abby's. Through this exchange, Abby learns that Paderewski is a musician.

Not only does Abby gossip with Gabby, she also gossips with me. Through this exchange, I also gain a concept of Paderewski (again, thanks to the anti-individualism thesis). I learn, through this exchange, that Paderewski is a politician.

Finally, Gabby comes to me with her gossip. She gossips about Paderewski, and I learn through this exchange that Paderewski is a musician.

In short, I have two pathways of information going back to Paderewski. The first one goes directly through Abby. The second one goes indirectly through Gabby and then to Abby. The directed graph is a triangle that looks like this:



What I don't realize is that Abby and Gabby were each gossiping to me about the same person on the two occasions. Nonetheless, I now have two premises at my disposal:

(1) Paderewski is a musician
(2) Paderewski is a politician

Given the anti-individuationalism thesis, the two tokens of <Paderewski> in the premises are tokens *of the same concept*. However, common intuition is that it would be *irrational* for me to draw the conclusion:

(C) There is someone who is both a musician and politician.

After all, *I didn't know* that Gabby and Abby were talking about the same person. (Nor is it irrational for me to fail to realize this. The information is coming from two sources without any reason to suspect a common original source.)

What follows from this? It follows that there's a conflict between two desiderata for concept individuation. On the one hand, we want concepts to be individuated according to the anti-individualism thesis. Or at least, most of us do—I think I speak for the majority here. Few of us would want to throw back to the old empiricist picture whereby concepts are formed *by the individual* on the exclusive basis of their own private experiences. In that old picture, concepts lose much of their explanatory power for communication, agreement / disagreement, and so on.

On the other hand, the Fregean wants concepts (i.e. thought constituents) to explain rational coordination. However, Kit Fine uses examples like the one above to argue that rational coordination between tokens is not an *equivalence relation*.

To make this argument, we first observe that in the context of my conversation with Gabby, owing to my deference, my tokens of <Paderewski> are rationally coordinated with Gabby's. That is to say that I can infer that someone is both F and G whenever Gabby tells me that <Paderewski is F; Paderwski is G>. Secondly, in Gabby's conversation with Abby, Gabby's tokens of <Paderewski> are coordinated with Abby's.

If the rational coordination of concept tokens formed an equivalence relation, it would follow that my tokens of <Paderewski> that defer to Gabby's will be rationally coordinated with Abby's. *But they're not*. They're not because when I token <Paderewski> in deference to Abby in the context of my conversation with her, those tokens are *not* rationally coordinated with my tokens of <Paderewski> that defer to Gabby. (We may assume, for the sake of argument (or reductio) that all of Abby's tokens are rationally coordinated with each other; so that I have two deferential channels going back to a concept source that is rationally integrated.)

If this argument is sound (and it's complicated enough for me to hedge with the conditional), then it would be a serious blow to the kind of Fregean discussed here. If rational coordination is not an equivalence relation, then there's no possible analysis of the form:

$<a_1>$ is rationally coordinated with $<a_2>$ iff they share the same ____.

Not the same sense, not the same syntax, not the same origin, not the same *anything*.

Conversely, if rational coordination *were* an equivalence relation, then we could define Fregean senses by reifying over equivalence classes of rationally coordinated tokens.[4] But if Fine's argument is sound, then we can't. So, as a result, the score is now 1-0 for the Russellians. (The earlier point for the Fregeans has been deducted.)

9. Is anti-individualism to blame?

Admittedly, all of this rests on the anti-individualist thesis. As I said, I think that *most* people will buy into

[4] Although I don't think that this move would be very Fregean in spirit. Senses are supposed to *explain* rational coordination; not the other way around.

it. But perhaps not everyone will. Is it possible to simply reject it, adhere to an individualistic view of concepts, and maintain that rational coordination between tokens *within* a single subject's cognition is an equivalence relation?

Even here, I think that there's room for doubt. I bet that we can transpose the problem into the cognitive system of a single individual. Let "Abby" be your perceptual system, "Gabby" be your memory storage and retrieval system, and "me" be the most central processing space of your cognitive system, where personal-level representations, like beliefs, become conscious (your global workspace).

Can there be failures of transitivity when representations get passed through these three nodes? It feels like if such failures exist in the interpersonal case then they exist in the intrapersonal case. Admittedly, we might recoil at the thought of "rational" coordination between tokens in a *sub-personal* system, like the perceptual system. It's *people* who are rational or irrational, not sub-systems. But the whole marriage between Fregeanism and RTM presupposed that we can speak of tokens processed in sub-personal systems as being rationally coordinated. Again, the idea was to treat Fregean senses as representations (types) in the RTM. But the life of a representation in the RTM is mostly subpersonal. These representations go in and out of the storage center, get passed along to the inference center, and then back to the storage center, and so on. It doesn't make much sense to say that *I* perform these computational processes. It's something that *my brain does.*

Perhaps therein lies the problem. Senses were originally introduced to explain when a *subject* (a person) is being rational or irrational. The original Frege railed against psychologism. Senses were impersonal abstract objects that constituted platonic propositions. Frege would roll over in his grave if he saw what the senses are being asked to do now. They're being treated as the medium over which the brain does its thing.

But just as an intestine cannot be rational for digesting fiber and irrational for rejecting lactose, a brain cannot be rational for integrating tokens of <Hesperus> and irrational for integrating tokens of <Hesperus> and <Phosphorus> (absent the subject's knowledge that *Hesperus is Phosphorus*). Obviously there has to be *some* story to tell about the interplay between personal-level intentional states and processes (belief, thought, inference) and sub-personal cognitive states and processes. But it's beyond my paygrade.

10. Conclusion

Where does this all leave us? Look, I don't take myself to have *proven* that there's no such thing as a Fregean Fodorian (who wears a fedora). What I have done, however, is present an obstacle to the hope that the philosophy of mind will track our epistemological concerns—something that Fregeans are most wont to hope. The "where do we go from here?" question then looks to splinter off into so many questions that the issue becomes well nigh intractable. So I don't really know what to say next.

Appendix

Perhaps I wrote this too soon. Recanati's combination of Fregeanism with RTM appeared to be initially motivated by rationality considerations, but at the end of the book I learned that the real concern is nothing of the kind. According to him, the real difference between an RTM that implements a one-level

semantic theory (Russellianism) and a two-level Fregean version is whether the thing that plays the vehicle role has any further semantic functions besides referring. If there's more than one distinct semantic function, then it's a two-level semantics, and therefore it's Fregean.

So what's the additional role? According to Recanati, it's the mental analogue of Kaplanian character. Suppose I think "I am tired" and you also think "I am tired". Uncontroversially, you and I think distinct thoughts, since my thought is about me and your thought is about you. Nonetheless, Recanati wants to say that, on some level, the representations we employ share the same "character". My representation <I> has the semantic *function* of referring to the subject employing it, and your representation <I> has the same semantic *function* of referring to the subject employing it. Hence, there's a semantic function they both share. And since it's a *semantic* function, theorizing about it requires a two-level semantic theory.

There are two questions to ask here. First, what is the cash value of calling the thing that plays the Kaplanian character role "semantic" (as opposed to pre-semantic, meta-semantic, pragmatic, functional or whatever)? Secondly, is the thing that plays the Kaplanian character role worthy of being called "sense"?

To the first question, at one point Recanati says that it's a mere terminological issue. Nothing hangs on whether we call Kaplanian character "semantic" or "metasemantic". I take it that the cash value of calling it 'semantic' in the linguistic case has to do with the fact that characters are conferred on expression types by community wide linguistic conventions—and that's what semantics is all about. But mental representation types—in the sense that your <I> and my <I> is a common functional type, despite their implementation being implementations of distinct concepts—are not governed by conventions. Rather, their functions are conferred by… by what? Evolution? God? I don't really know how to think about this, and besides, I don't really want to. Suffice to say the analogy between the linguistic case and the mental case breaks down.[5]

Is Kaplanian character at all like sense? There's probably a host of reasons against this. But most of all, it just isn't natural to say there's a sense in which you and I "think the same thought" when I think of myself as tired and you think of yourself as tired. The 2D semanticist wants us to recognize that *on some level* "the same thought" is being thought. And maybe there's some theoretical advantage to this. But again, it's beyond my paygrade to comment on it here.

Works Cited

Evans, Gareth (1982). *The Varieties of Reference.* Ed. by John McDowell. Oxford University Press.

Fine, K. (2009). *Semantic relationism*. John Wiley & Sons.

Frege, Gottlob. (1948) 'Sense and Reference', *The Philosophical Review*, 57/3: 209–30.

Fodor, Jerry. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

[5] A long time ago, I wrote a paper "What is analyticity" that took a two-dimensional approach to explaining analyticity along Kaplanian lines. It was a naive paper, and I have to rethink a lot of it. But one thing that surprised me about it is that, if we take the *Kaplanian* approach to explaining analyticity for sentences, then we get no corresponding notion of 'conceptual truth' for thoughts. The theory really only applies to linguistic items. Well, maybe I'd have to rethink this. The problem is that I just don't really have a good handle on indexicality for *mental* representations.

Fodor, Jerry. (1995). *The elm and the expert: Mentalese and its semantics*. MIT press.

Gray, Aidan. “Thinking the Same-ish” *Sharing Thought,* José Bermudez, Matheus Valente, Víctor M. Verdejo (eds) . Forthcoming.

Ichikawa, J. J., & Jarvis, B. W. (2013). *The rules of thought.* Oxford University Press.

Kaplan, D. (1989b) ‘Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals’, in J. Almog, J. Perry & H. Wettstein (eds.) *Themes From Kaplan*, 581–63. Oxford: OUP.

Kripke, Saul (1979). “A puzzle about belief”.

Recanati, François. (2012). *Mental files*. Oxford University Press.

Sainsbury, R. M., & Tye, M. (2012). *Seven puzzles of thought: And how to solve them: an originalist theory of concepts*. Oxford University Press.

Salmon, Nathan (1986). *Frege's puzzle*. Ridgeview Publishing Co.

Shea, N. (2018) *Representation in Cognitive Science.* Oxford: OUP.

Taylor, K. A. (2019) *Meaning Diminished: Toward Metaphysically Modest Semantics.* Oxford: OUP.