

Abstract: Representing Reason

Nicholas Koziolk
University of Chicago

May 27, 2015

1 Introduction

My dissertation is about our ability to represent one another's thoughts. It is about that capacity, or set of capacities, in virtue of which I can represent you as knowing that the earth revolves around the sun; as believing that *The Brothers Karamazov* is more profound than *War and Peace*; as assuming (or at least hoping) that this abstract will be informative; and so on. It is also, albeit only indirectly, about our ability to represent one another's reasoning. It is about that capacity, or set of capacities, in virtue of which I can represent you not just as (for example) believing something, but as believing it, and believing it reasonably, precisely because you believe something else. Since, in the fundamental case, such representations of your thoughts and reasoning are *judgments* about your thoughts and reasoning, the dissertation is, most fundamentally, about a certain subcapacity of the capacity to judge: namely, the capacity to judge that another thinker thinks some particular thing or reasons in some particular way.

Many philosophers will know my topic by a different name. To speak in one familiar philosophical idiom, then, my dissertation is about propositional attitude ascriptions, or attributions. But I write, instead, of our ability to represent one another's thoughts and reasoning because, on the view I defend in the dissertation, this ability is what ultimately underlies and grounds our use of propositional attitude ascriptions, i.e., of sentences like 'Galileo believes that the earth revolves around the sun'.

The dissertation is, and was conceived as, part of a larger project in the metaphysics of mind, and specifically in that part of the metaphysics

of mind that was once called the theory of judgment and, more recently, the theory of thought. The larger project is meant to issue, ultimately, in a representation of the whole of the faculty of reason. In the dissertation, however, my aim is only to produce a representation of one of the central components of that faculty: our ability to represent one another's thoughts – which ability is intimately bound up with our ability to represent one another's reasoning.

2 Aim

My central aim in the dissertation, to speak in something like the familiar philosophical idiom mentioned above, is to develop and defend an account of thoughts that ascribe propositional attitudes, thoughts of the form $\langle S \phi$ that $p \rangle$ – for example, the thought that Plato believes that Socrates is wise. For it is in thinking such thoughts (most fundamentally, in judging them to be true) that we represent one another's thoughts; and so it is in thinking such thoughts that we represent one another's reasoning. The account I will defend derives from – though, importantly, it is not identical with – the theory of indirect discourse famously proposed by Gottlob Frege (1892: 28/153–54). My account can be characterized in terms of two of its central theses. The first is that the truth-conditions of the thought that $S \phi$ that p depend on the thought that p . The second is that the thought that $S \phi$ that p has the thought that p as a component part. Since Frege himself accepts the first of these theses, but rejects the second, the result of combining them is a neo-Fregean theory of thoughts that ascribe propositional attitudes – a neo-Fregean theory of ascription, as I call it.

The reasons for accepting the first thesis are familiar. Specifically, there are familiar reasons for thinking that the truth-conditions of the thought that $S \phi$ that p cannot depend on anything less finely individuated than a Fregean thought. For example, 'Mark Twain' was the pen name of one Samuel Clemens. But not everyone knows this fact. So some people believe that Mark Twain wrote *The Adventures of Huckleberry Finn*, but do not believe, and would probably even deny, that Samuel Clemens wrote *The Adventures of Huckleberry Finn* (after all, they know perfectly well that Mark Twain wrote it). So the thought that Mark Twain wrote *Huck Finn* and the thought that Samuel Clemens wrote *Huck Finn* have different truth-conditions. But the only way to guarantee that these thoughts have

different truth-conditions is to maintain that the truth-conditions of the thought that $S \phi$ s that p depend on an “entity” (associated with the thought-component $\langle p \rangle$) that is individuated *in terms of* the propositional attitudes themselves: we need a view on which the “entity” associated with $\langle p \rangle$ is distinct from the “entity” associated with $\langle q \rangle$ *if* (though perhaps not only *if*), say, the *belief* that p is distinct from the *belief* that q . But – at least according to an influential interpretation proposed by Gareth Evans (1982: 18–19) – this is precisely how Fregean thoughts are individuated. So, I content, Fregeanism represents the only genuinely plausible approach to the semantics of thoughts that ascribe propositional attitudes. (I explain these points in detail §1.4 of the dissertation.)

The reasons for accepting the second thesis are less familiar. So it will be worth going into a bit more detail about this thesis here, especially since it is the second thesis that is distinctive of the neo-Fregean theory of ascription I defend in the dissertation.

3 Motivation

The second thesis of the neo-Fregean theory of ascription can be motivated by way of the following considerations.

Human beings are rational animals. Because we are rational, we are able to engage in self-conscious reasoning about the world. If I believe, for example, both that Socrates is a philosopher and that all philosophers are wise, I can reason from these premises to the novel conclusion that Socrates is wise. Equally, if I believe both that Socrates is a philosopher and that all philosophers are wise, and I acquire evidence that Socrates is not in fact wise, I can reason from this evidence to the novel conclusion that, contrary to my prior beliefs, either some philosophers are not wise or Socrates is not a philosopher.

We are also, however, and again because we are rational, able to engage in self-conscious reasoning about one another’s self-conscious reasoning. If I believe, for example, both that Plato believes that Socrates is a philosopher and that he believes that all philosophers are wise, I can reason from these premises to the novel conclusion that he (probably) believes that Socrates is wise. Equally, if I believe both that Plato believes that Socrates is a philosopher and that he believes that all philosophers are wise, and I acquire evidence that he does not in fact believe that Socrates is wise, I can

reason from this evidence to the novel conclusion that, contrary to my prior beliefs, (probably) either he does not believe that Socrates is a philosopher or he does not believe that all philosophers are wise.

How, we might ask, is our ability to engage in self-conscious reasoning about the world related to our ability to engage in self-conscious reasoning about one another's self-conscious reasoning?

A striking fact about the above examples – a fact that suggests that the question just posed is indeed a question worth asking – is that there is a close structural parallel between my reasoning about Socrates and my reasoning about Plato's reasoning about Socrates. In the first case, I reason from the premises that Socrates is a philosopher and that all philosophers are wise to the conclusion that Socrates is wise. In the second, I reason from the premises that Plato *believes* that Socrates is a philosopher and that Plato *believes* that all philosophers are wise to the conclusion that Plato (probably) *believes* that Socrates is wise. So suppose we represent my reasoning, in each case, as an inference. Then we have both:

- (1) Socrates is a philosopher
All philosophers are wise
Therefore, Socrates is wise

and:

- (2) Plato believes that Socrates is a philosopher
Plato believes that all philosophers are wise
Therefore, Plato (probably) believes that Socrates is wise.

Here the structural parallel between my reasoning in each case is rendered visible in the fact that the sentences 'Socrates is a philosopher', 'All philosophers are wise', and 'Socrates is wise' appear both in our representation of my reasoning about Socrates and in our representation of my reasoning about Plato's reasoning about Socrates; and they appear, moreover, in the very same order.

There is also, of course, an obvious and important difference between these two inferences, a difference already marked by my inclusion of the parenthetical 'probably' in the third line of the second. The difference is that the first inference is deductively valid, while the second is at best inductively strong. Still – and this is the crucial point – even if the second inference is one that is merely inductively strong, it is plausible that

its strength is at least partly a function of the deductive validity of the inference schema:

- (3) a is F
 Everything F is G
 Therefore, a is G ,

a schema that, it seems, is instantiated – though, apparently, in rather different ways (or, more literally, in different places) – in both of the above inferences.

There is, in fact, positive reason for thinking that schema (3) is indeed instantiated in inference (2). Thus, consider the following two ways of schematizing inference (2):

- (4) S ϕ s that q
 S ϕ s that r
 Therefore, S (probably) ϕ s that p

and:

- (5) S ϕ s that a is F
 S ϕ s that everything F is G
 Therefore, S (probably) ϕ s that a is G .

Clearly, instances of the more abstract schema (4) are not guaranteed to be inductively strong. Here, for example, is a perfectly fine instance of (4) that would make for a terrible inference: Martin believes that grass is green; Martin believes that snow is white; therefore, Martin believes that dirt is grass. By contrast, every instance of the more determinate schema (5) is guaranteed to be just as strong as inference (2). So it is the more determinate schema (5), and not the more abstract schema (4), that gives the form of inference (2).

The upshot is that the form of the thought that Plato believes that Socrates is wise, for example, is given, not by the abstract schema:

- (6) S ϕ s that p ,

but, rather, by the more determinate schema:

- (7) S ϕ s that a is F .

In other words, to generalize: the form of the thought that $S \phi$ s that p is partially characterized by the form of the thought that p : the internal structure of the thought that p is also part of the internal structure of the thought that $S \phi$ s that p . But the only way to account for this fact – or so I contend – is to show that the thought that p is a component *part* of the thought that $S \phi$ s that p . It is not enough, for example, to say that the thought-component \langle that p \rangle is *a way of thinking of* the thought that p . Rather, the thought-component \langle that p \rangle must *bear the form of* the thought that p . And it is hard to see how it can do that without *being* the thought that p . (I explain these points in detail in §1.5 of the dissertation.)

4 Summary of the dissertation

Following the introductory Chapter 1, the dissertation is divided into two parts. Part I treats of the general theory of thought (as I call it), a theory to which the subsequent account of thoughts that ascribe propositional attitudes – the theory of ascription – is thus answerable. Part II then explains and defends the neo-Fregean theory of ascription.

Part I: The Theory of Thought

Chapter 2: Capacities to Judge. The central task of Chapter 2 is to explain and defend an account of thoughts that respects Gareth Evans's (1982: 18–19) Intuitive Criterion of Difference:

the Intuitive Criterion of Difference: If, for some attitude ϕ , it is possible for someone to ϕ that p without ϕ ing that q , then the thought that p is distinct from the thought that q .

I defend the Criterion by providing an account – in fact, a definition – of thoughts in terms of a particular propositional attitude: namely, judgment. Specifically, I define a thought as a certain sort of capacity to judge: an Aristotelian first potentiality. Correlatively, I define a belief as a related but different sort of capacity to judge: an Aristotelian second potentiality.

As I understand these ideas, a first potentiality is a capacity the actualization of which is itself a capacity to perform some particular act. So, for example, the capacity to speak French, understood as a first potentiality, is something possessed by every normal human being; it is, as we might

put it, the capacity to *learn* (to speak) French. The capacity to *speak* French, understood as a second potentiality (or first actualization), on the other hand, is something possessed only by actual French speakers. It is a capacity acquired when one *learns* French. So, importantly, the capacity to *speak* French (a second potentiality) is itself the *actualization* of the capacity to *learn* (to speak) French (a first potentiality). And the capacity to speak French is itself actualized in actually *speaking* French. But, crucially, there is really just *one* capacity here, with two stages of actualization. The capacity is the capacity to speak French, which is actualized, first, in learning to speak French, and is actualized, second, and most fully, in actually speaking French.¹

In terms of this structure – the structure of a capacity with first and second actualizations – we can, given an appropriate act or activity, define two different capacities, the first and second capacities to perform the relevant act or activity. My account of thoughts, then, involves the application of this idea to the act of *judgment*. Thus, the thought that Socrates is wise, for example, is the (first) capacity to *judge* that Socrates is wise. To possess this capacity is, quite simply, to possess its component concepts (also capacities): the relevant way of thinking of Socrates (namely, as Socrates) and the concept of being wise. The *belief* that Socrates is wise is then the first actualization of this capacity, an actualization achieved through the acquisition of evidence to the effect that Socrates is wise. The actualization of the belief that Socrates is wise is then, finally, the judgment that Socrates is wise. So the belief is *exercised* or *actualized* in, for example, reasoning with it, where such reasoning consists, in part, in *judging* that Socrates is wise.

The upshot is that a thought is a capacity. In fact, a thought is a *complex* capacity: it is a capacity made of up component capacities, capacities that are exercised together in judging what the thought is a capacity to judge. In these terms, the central question of the dissertation is one about the component capacities of, for example, the thought that Plato believes that Socrates is wise. The central thesis (the second thesis) of the neo-Fregean theory of ascription is that the thought – the capacity to judge – that Socrates is wise is a component part of the thought – the capacity to judge – that Plato *believes* that Socrates is wise. Thus, the very same capacities that would be exercised in judging that Socrates is wise are also exercised in

¹I am indebted here to Kosman 2013: 57–62.

judging that Plato believes that Socrates is wise.

Chapter 3: Sense, Reference, and Semantic Value. In Chapter 3, I go on to introduce the philosophical and terminological framework in which the neo-Fregean theory of thought – and so the neo-Fregean theory of ascription – is to be developed. Here I motivate and explain a threefold distinction between sense, reference, and semantic value and relate it to Frege’s familiar two-fold distinction between sense and *Bedeutung*. I show, in particular, that the concept of semantic value can be explained without appeal to Frege’s notion of sense: the semantic value of an expression is simply the contribution it makes to the truth-conditions of sentences in which it occurs. I also argue, drawing on the conclusions of Chapter 2, that the concepts of sense and reference, and Frege’s famous distinction, can be explained without appeal to the notion of semantic value: the sense of a sentence is a thought, and the sense of a subsentential expression is an appropriate *concept*, that is, an appropriate subcapacity of the complex capacity – the thought – that is the sense of the sentence; and the reference of an expression is an entity associated with the capacity that serves as the expression’s sense (on the assumptions I make in the dissertation, it is the object, function, or truth-value of which the associated sense is a way of thinking).

With that work done, I claim, it is easy to see that Frege connects the concepts of sense and reference to the concept of semantic value in a particular way, by employing the term ‘*Bedeutung*’ to express *both* the concept of reference *and* the concept of semantic value. Since these concepts are distinct and independent of one another, however, it becomes apparent that the connection is questionable. This point is crucial to the work of the dissertation, because it turns out that the neo-Fregean theory is simply indefensible so long as the relevant connection between the concepts of sense and reference and the concept of semantic value remains in place. An important goal of Part II, then, is to make good on the promise of Chapter 3, and thus to show that the connection can be abandoned without doing undue violence to Frege’s core theory of sense and reference.

Part II: The Theory of Ascription

Chapter 4: Propositional Ascriptions. Thus, I turn, in Part II, to the theory of ascription. I begin in Chapter 4 by presenting, in some detail, both the classical Fregean and the neo-Fregean theories of ascription, as well as a third, “two-level” theory, employing the philosophical and terminological framework of Chapter 3 throughout. By presenting the three views in the same – and, I claim, a neutral – terminology, we are better able to compare their respective advantages and disadvantages.

The central difference between the theories is this: On the classical theory, each expression of a language needs to be assigned not only a sense and a reference, but also an indirect sense and reference, a doubly indirect sense and reference, and so on), to be associated with the expression under each iteration of the operator ‘ $S \phi$ that ...’. The two-level theory, by contrast, assigns each expression, in addition to its “customary” sense and reference, an indirect sense and reference, to be associated with the expression whenever it occurs embedded under the operator ‘ $S \phi$ s that ...’, no matter how many times it is so embedded. The neo-Fregean theory, however, assigns each expression *only* a single sense and reference. Thus, the application of the operator ‘ $S \phi$ s that ...’ has, on this view, no affect on the sense or reference of any expression. Importantly, however, the theories agree about the truth-conditions of sentences that ascribe propositional attitudes (even when the attitude operator is iterated). The differences in assignments of sense and reference – which would otherwise make for differences in truth-conditions – is made up for by corresponding differences in compositional rules.

I close the chapter with a discussion of one particular advantage of the neo-Fregean over the classical Fregean and two-level views: the first, unlike the other two, I argue, is semantically innocent in the sense made famous by Donald Davidson (1968: 108) – that is, it does not require us to allow that the same expression has a different sense and reference in different linguistic contexts. That, I suggest, gives us a first reason to reject the connection Frege forges between the concepts of sense and reference and the concept of semantic value.

Chapter 5: Canonical Ways of Thinking. A well-known objection to the classical Fregean theory of ascription has it that any language correctly

describable by the theory would have to be unlearnable.² Recently, however, a number of philosophers³ have suggested that we can answer this objection by introducing a rule that determines the indirect sense of an expression on the basis of its customary sense, its doubly indirect sense on the basis of its singly indirect sense, and so on, all the way up the infamous infinite hierarchy of senses. In Chapter 5, by way of providing some additional motivation for the neo-Fregean theory of ascription, I show that this attempt to climb Frege's hierarchy fails. As I explain there, our understanding of the proposed rule turns out to reside entirely in the very thing the rule was introduced to explain: our grasp of thoughts that ascribe propositional attitudes. I conclude that, in order to save Frege, we must save him from himself. We must reject a key, and so far unquestioned, feature of his formal framework: namely, his extensionalism, the view that the truth-value of a sentence is a function (only) of the references of its parts. It is this feature of his formal framework, I argue, that forges the important connection, mentioned above, between the concepts of sense and reference and the concept of semantic value.

Chapter 6: Extensionality and the Composition of Thoughts. In Chapter 6, I continue my defense of the neo-Fregean theory of ascription by responding to some influential objections, originally due to Tyler Burge (1979, 2004). In so responding, I sketch a novel conception of logical form, according to which the logical form of a sentence (or a thought) is – to put it roughly – *the way in which* its truth-conditions depend on the semantic features of its parts. I then apply that conception of logical form to thoughts and sentences that ascribe propositional attitudes, with the result that there are two distinct ways in which a referential or predicative capacity (the basic sub-capacities of thoughts, introduced in Chapter 2) can occur in a thought. This idea, it proves, suffices to undermine the most important objection to the neo-Fregean view, which objection turns out to rest solely on the undefended assumption that the truth-conditions of a thought can never depend on the parts of the thought themselves, but must instead depend only on the objects and functions of which its parts are ways of thinking.

²The objection is Davidson's (1965).

³Burge (1979 and 2004), Kripke (2008), Peacocke (2009), and Parsons (2009).

Chapter 7: Predicative Ascriptions. In the final chapter, Chapter 7, I argue that the neo-Fregean view, as presented in Chapters 4 and 5, can be extended to provide an account of *de re* and *de se* ascriptions – ascriptions of the forms ‘*S* ϕ s of *a* that it is *F*’ and ‘*S* ϕ s that she herself is *F*’, respectively. The central problem facing the neo-Fregean theory of ascription, in this area, concerns the semantic role of the pronouns ‘it’ and ‘she herself’: in short, the neo-Fregean seems to be forced to choose between the view that the semantic values of the pronouns are the *references* of their antecedents and the view that their semantic values are the *senses* of their antecedents; but neither of these options is acceptable. I show, however, that the dilemma is false. We can escape it by refusing to assign any semantic values at all to the pronouns in *de re* and *de se* ascriptions. What seems to require such an assignment is the assumption that *de re* and *de se* ascriptions, like *de dicto* ascriptions (i.e., those of the form ‘*S* ϕ s that *a* is *F*’), ascribe complete thoughts, but thoughts of a special kind. I argue that we should reject this assumption. Instead, I claim, we must see *de re* and *de se* ascriptions as ascriptions of *predications* (ascriptions of *self*-predications, in the case of *de se* ascriptions), rather than ascriptions of complete thoughts. The result is that both *de re* and *de se* ascriptions are to be recognized as primitive forms of thought. I show that these forms of thought can be explained in conformity with the neo-Fregean theory of ascription developed in Chapters 4–6, and that the resulting theory of *de re* and *de se* ascriptions gives the right results.

Conclusion

Throughout, what the neo-Fregean view exploits is the idea that we represent the thoughts of others through our own representations of the world. Wherever the same representational feature shows up in different thoughts, the same singular term or predicate shows up in adequate regimentations or formalizations of the sentences that express those thoughts. The neo-Fregean view exploits this insight in order to earn us the right to the view that each of the following thoughts involves the concept (in the usual philosophical sense) of being *F*: the thought that *a* is *F*, the thought that *S* ϕ s that *a* is *F*, the thought that *S* ϕ s of *a* that it is *F*, and the thought that *S* ϕ s that she herself is *F*. It thus earns us the right to claim that the relations we see to obtain between these thoughts – relations reflected in our use of the letter ‘*F*’ throughout – are *logical* relations. In this way, it earns us

the right to claim that what we have here are representations of the logical forms of these thoughts. It thus allows us to do justice to the view of the relation between self-conscious reasoning and self-conscious reasoning about reasoning that was described, briefly, above.

References

- Almog, Joseph and Paolo Leonardi (editors). 2009. *The Philosophy of David Kaplan*. Oxford: Oxford University Press.
- Beaney, Michael (editor). 1997. *The Frege Reader*. Malden, MA: Blackwell.
- Burge, Tyler. 1979. Frege and the hierarchy. *Synthese* 40(2), 265–281. Reprinted in Burge 2005, 153–166. References give the page number of the reprint.
- 2004. Postscript to ‘Frege and the hierarchy’. In Burge 2005, 167–210.
- 2005. *Truth, Thought, Reason*. Oxford: Oxford University Press.
- Davidson, Donald. 1965. Theories of meaning and learnable languages. In Davidson 2001, 3–15.
- 1968. On saying that. *Synthese* 19, 130–146. Reprinted in Davidson 2001, 171–183. References give the page number of the reprint.
- 2001. *Inquiries Into Truth and Interpretation*. Oxford: Oxford University Press.
- Evans, Gareth. 1982. *The Varieties of Reference*. Edited by John McDowell. Oxford: Oxford University Press.
- Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100, 25–50. Translated by Max Black and reprinted as “On Sinn and Bedeutung” in Beaney 1997, 151–171. References give the page number of the original, then the page number of the reprint.
- Kosman, Aryeh. 2013. *The Activity of Being*. Cambridge, MA: Harvard University Press.
- Kripke, Saul. 2008. Frege’s theory of sense and reference: some exegetical notes. In Kripke 2011, 254–291.
- 2011. *Philosophical Troubles*. Oxford: Oxford University Press.

Parsons, Terence D. 2009. Higher-order senses. In Almog and Leonardi 2009, 45–59.

Peacocke, Christopher. 2009. Frege’s hierarchy: a puzzle. In Almog and Leonardi 2009, 159–186.