

SELF-VERIFICATION AND THE CONTENT OF THOUGHT<sup>1</sup>

Descartes famously argued, on purely conceptual grounds, that even an extremely powerful being could not trick him into mistakenly judging that he was thinking. Of course, it is not necessarily true that Descartes is thinking. Still, Descartes claimed, it is necessarily true that if a person judges that she is thinking, that person is thinking. Following Tyler Burge (1988) we call such judgments ‘self-verifying.’ More exactly, a judgment *j* performed by a subject *S* at a time *t* is self-verifying if and only if the fact that *S* has made *j* at *t* entails that *j* is true at *t*.<sup>2</sup> Burge follows Descartes in claiming that the category of conceptually self-verifying judgments includes (but is not restricted to) judgments that give rise to sincere assertions of sentences of the form, ‘I am thinking that *p*’. We call such judgments ‘cogito’ judgments.

In this paper I argue that Burge’s Cartesian insight is hard to reconcile with Fregean accounts of the content of thought. Theorists have tried to account for the self-verifying status of cogito judgments by arguing that the second-order judgment that one is thinking that *p* *contains* the thought that *p* as a part of it. I argue (§1) that the accuracy of the containment model is entailed by a Russellian view of content (according to which “belief” contexts are both extensional and transparent) when the Russellian view is attached to some fairly uncontroversial assumptions. The accuracy of the containment model is also entailed by non-hierarchical Fregean views of content according to which expressions in oblique contexts both denote and express first-level senses (§2), but there are compelling arguments against non-hierarchical Fregean views (§3). Moreover, the extremely plausible account of self-verification provided by Russellianism is not entailed by theories that follow Frege in accepting a hierarchy of senses and so allow that expressions in oblique contexts express senses that are not identical to the senses they there denote (§4). If these arguments are sound, Fregeans must accept the hierarchy of senses, and to accept the hierarchy they must substantially augment their view to account for self-verification. I go on to explain how a Fodorian account of senses might try to take up this challenge, but I

show that the Fodorian solution must posit conceptually necessary connections between distinct concreta (§5). Finally (§6) I consider Alonzo Church's (1974) description of 'standard' higher-level senses and show why some notion of this kind must be invoked by the Fregean to account for self-verification. But the mere existence of standard higher-level senses is not up to the task, even if we follow Church in positing logical relations between standard senses and the lower-level senses they determine. Fregeans must go further than Church and claim that standard higher-level senses have a particular composite structure. It seems, then, that Burge's intuitively compelling claim that cogito judgments are conceptually self-verifying poses a real challenge to neo-Fregean theories of content.

### 1. A RUSSELLIAN ACCOUNT OF SELF-VERIFICATION

Consider *S*'s judgment that she is thinking that John is nice. Burge joins Descartes in claiming that necessarily, if *S* judges that she is thinking that John is nice, then it is true that *S* is thinking that John is nice. We can establish this conclusion from four premises.

- (1) Judgment is a subspecies of cognitive consideration: to judge that *p* one must consider the proposition that *p*.
- (2) When a subject considers a content, she therein considers its constituents. (For example, to consider the proposition that  $\beta$  is *F* is to both consider  $\beta$  and consider *F*.)
- (3) Considering the proposition that *p* is a way of thinking that *p*; to consider the proposition that *p* is to entertain the proposition that *p*. (For example, considering the proposition that  $\beta$  is *F* is a way of thinking that  $\beta$  is *F*.)
- (4) The content of a subject's higher-order judgment that she is thinking that *p* literally includes the constituents of *p*. (For example, the content of *S*'s judgment that she is thinking that  $\beta$  is *F* literally includes the constituents of the proposition that  $\beta$  is *F*.)

The argument runs as follows. First, we establish premise (4) by assuming a 'neo-Russellian' account of content. Let ' $\beta$ ' be a proper name and '*F*' a predicate of a natural language. Suppose we say that the judgment that  $\beta$  is *F* is a relation between a subject and a proposition constituted by  $\beta$  and *F*:  $\langle \beta, F \rangle$ . We can then say that for *S* to judge that she is thinking that  $\beta$  is *F* is for her to be

related by judgment to  $\langle S, \text{thinking}, \langle \beta, F \rangle \rangle$ . According to premise (1), judgment necessarily involves cognitive consideration; so in being related by judgment to  $\langle S, \text{thinking}, \langle \beta, F \rangle \rangle$ ,  $S$  is related by consideration to  $\langle S, \text{thinking}, \langle \beta, F \rangle \rangle$ . If premise (2) is correct, then in being related by consideration to  $\langle S, \text{thinking}, \langle \beta, F \rangle \rangle$ ,  $S$  is related by consideration to its constituents, which include (among other things)  $\langle \beta, F \rangle$ . And if premise (3) is correct, considering  $\langle \beta, F \rangle$  is a way of thinking that proposition. Thus, judging that one is thinking that  $\beta$  is  $F$  conceptually entails thinking that  $\beta$  is  $F$ . Informally (and using a particular example): To judge that one is thinking that John is nice, one must consider the proposition that one is thinking that John is nice. To consider the proposition that one is thinking that John is nice, one must consider oneself, thought, and the proposition that John is nice. Considering the proposition that John is nice is a way of thinking that John is nice. Thus, the fact that one is thinking that John is nice is entailed by the fact that one judges that one is thinking that John is nice.<sup>3</sup>

What can be said on behalf of (1)–(4)? It seems as though (1) and (3) are uncontroversial when taken in a certain way. That is, there is a sense (or use) of ‘judgment’ according to which judging-true a proposition means both considering that proposition and endorsing or assenting to it. There is a sense of ‘thinking’ according to which entertaining or considering a proposition that one might (or might not) judge-true means engaging in a kind of thought. So as long as we remain clear that the defender of conceptual self-verification is limiting her claim to instances in which one considers and endorses a proposition, and is claiming that, in such instances, one must be thinking that proposition in this very general sense, (1) and (3) seem unobjectionable. Similarly, there is a sense in which considering a proposition requires considering its constituents: one does not count as considering the proposition that John is nice unless one has, in some very general way, considered John and niceness. So interpreted, (2) is unobjectionable.<sup>4</sup>

Of course, (4) is much more problematic. As shown above, (4) follows from a neo-Russellian view of content, but many contemporary thinkers reject such a view. Indeed, Burge (1977) explicitly accepts a Fregean account of thought-content. So it remains to be seen whether Burge (and other neo-Fregeans) can reasonably accept (4) – an important premise in the argument for the conceptually self-verifying status of cogito judgments.

## 2. AN ATTEMPT AT A FREGEAN ACCOUNT OF SELF-VERIFICATION

How might a Fregean argue for the self-verifying status of a subject's judgment that she is thinking that  $p$ ? On the Fregean view, the first-order thought that John is nice is a relation between a subject and a proposition where the proposition is constituted by a sense or mode of presentation (MOP) determining John and an MOP determining niceness. Let  $\alpha$  range over natural language expressions, and let ' $\{\alpha\}$ ' denote the sense of  $\alpha$ . We then represent the purportedly Fregean content of the thought that John is nice as  $\langle\{\text{John}\}, \{\text{niceness}\}\rangle$ , where ' $\{\text{John}\}$ ' denotes the MOP expressed by the English expression 'John', ' $\{\text{niceness}\}$ ' denotes the MOP expressed by 'nice', and ' $\langle\{\text{John}\}, \{\text{niceness}\}\rangle$ ' denotes the proposition that John is nice.<sup>5</sup> According to the Fregean, then, in thinking that John is nice,  $S$  is related by thought to  $\langle\{\text{John}\}, \{\text{niceness}\}\rangle$ .

What, though, is the content of  $S$ 's introspective judgment that she is thinking that John is nice? If we accept premise (4) then  $S$ 's second-order judgment that she is thinking that John is nice is in part constituted out of the very MOPs that constitute her first-order thought:  $\{\text{John}\}$  and  $\{\text{niceness}\}$ . We might take this to suggest that  $S$ 's higher-order judgment is *not* constituted by a second-level MOP determining the first-level MOP expressed by 'John' (nor an MOP determining the MOP expressed by 'nice'). The content of  $S$ 's judgment that she is thinking that John is nice would then be  $\langle\{I\}, \{\text{thinking}\}, \langle\{\text{John}\}, \{\text{niceness}\}\rangle\rangle$ . If premise (1) is right, then to judge-true  $\langle\{I\}, \{\text{thinking}\}, \langle\{\text{John}\}, \{\text{niceness}\}\rangle\rangle$ ,  $S$  must consider or entertain  $\langle\{I\}, \{\text{thinking}\}, \langle\{\text{John}\}, \{\text{niceness}\}\rangle\rangle$ . If premise (2) is correct, then in considering  $\langle\{I\}, \{\text{thinking}\}, \langle\{\text{John}\}, \{\text{niceness}\}\rangle\rangle$ ,  $S$  must consider its constituents, one of which is  $\langle\{\text{John}\}, \{\text{niceness}\}\rangle$ . Finally, if premise (3) is correct, in considering  $\langle\{\text{John}\}, \{\text{niceness}\}\rangle$ ,  $S$  is thinking  $\langle\{\text{John}\}, \{\text{niceness}\}\rangle$ . Thus, if  $S$  judges that she is thinking that John is nice, she is thinking that John is nice.

Christopher Peacocke (1999) explicitly adopts this account of self-verification. He accepts a Fregean account of content, and his 'redployment claim' states that when an expression occurs in an embedded context it denotes the very MOP it there expresses:

The concepts (senses, modes of presentation) that feature in first-level thoughts not involving propositional attitudes are the very same concepts which feature in thoughts about someone's propositional attitudes . . . The sense of a word occurring in contexts

not involving propositional-attitude constructions is the same sense which is redeployed when the word occurs within the scope of propositional-attitude verbs. (1999: 246)

Such a view allows first-order senses, but rejects what is often called ‘the hierarchy of senses.’<sup>6</sup> According to Peacocke, “It is this which makes it possible for an intentional content in a contextually self-verifying self-ascription be, as Burge says, both thought and thought about in a single performance” (1999: 250).

### 3. FREGEAN COMMITMENT TO THE HIERARCHY

But if there are MOPs there must be senses of more than one level. We argue as follows.<sup>7</sup> Let ‘{Opus 132}’ denote the customary sense of ‘Opus 132’. And consider the false (indeed bizarre) proposition that {Opus 132} is often played with passion. Assume, for a reductio, that there is no hierarchy. ‘{Opus 132}’ has a sense. If it does not express a second-level MOP, it must express the very first-level MOP expressed by ‘Opus 132’. (What other sense could it have?) The proposition expressed by ‘{Opus 132} is often played with passion’ is therefore composed of the same constituents in the same manner of combination as is the proposition expressed by ‘Opus 132 is often played with passion’. So these propositions are the same. But the proposition that Opus 132 is often played with passion is true and (again) the proposition that {Opus 132} is often played with passion is false. So we cannot say that every MOP-denoting expression expresses the very MOP it denotes.

This argument assumes that senses can be denoted by expressions that occur outside of ordinary oblique contexts like those that are (arguably) created by propositional attitude verbs.<sup>8</sup> And it assumes for a reductio that these unembedded expressions behave in the very ways that, according to the non-hierarchical Fregean view, embedded expressions behave: they express the very senses they denote. No doubt, the argument shows that these assumptions lead to absurdity: if there are senses, there must be more than one level of senses. Still, a neo-Fregean might accept the existence of the hierarchy but deny that higher-level senses enter into the content of higher-order thoughts. According to this view, ‘{Opus 132}’ does indeed express a second-level sense, but ‘Opus 132’ as it occurs in ‘Bela believes Opus 132 is a masterpiece’ expresses the very first-level sense it there denotes.<sup>9</sup>

But arguments presented by Burge (1979) can be adapted to impugn even this view. Our argument assumes for a *reductio* that ‘Opus 132 is a masterpiece’ expresses the same sense whether it appears in an oblique or a transparent context, so that when it occurs alone ‘Opus 132 is a masterpiece’ expresses the same proposition that it expresses when embedded in ‘Bela believes that Opus 132 is a masterpiece’. To draw an absurd result from this assumption we follow Burge and introduce a formal expression, the extensional and transparent-context-maintaining operator ‘**T**’, the extension of which is a two-place relation between subjects and objects of belief. ‘**T**( $x$ ,  $y$ )’ is true if and only if ‘ $x$ ’ denotes a subject  $x$ , ‘ $y$ ’ denotes a proposition  $y$  and  $\langle x, y \rangle$  is a member of the extension of ‘**T**’. When ‘ $x$ ’ in this schema is replaced with a person’s name, and ‘ $y$ ’ is replaced with a sentence, the resulting sentence will be true iff the denoted person *believes* whatever it is that is denoted by the sentence. According to the truth-conditions supplied above, all such sentences will come out false, because sentences, if they denote at all, do not denote propositions. But ‘**T**’ can be used to express true propositions. Suppose that the proposition that Opus 132 is a masterpiece is the last proposition Kevin considered before going to bed. We introduce the name ‘Michael’ to denote this proposition. Then ‘**T**(Bela, Michael)’ is true just so long as Bela believes that Opus 132 is a masterpiece.

Now consider two Fregean theories about the behavior of English sentences in oblique contexts. The first says that when ‘believes that’ is attached to a sentence the denotation of the sentence shifts; whereas an unembedded sentence expresses a proposition and denotes a truth-value, when attached to ‘believes that’ a sentence will denote the proposition it customarily expresses. We now claim that despite the differences in behavior postulated between ‘**T**’ and the English expression ‘believes that’, we can stipulate that these expressions have the same sense. A formal expression can be given any sense we like unless the stipulation is incoherent; so what could prevent us from stipulating that ‘**T**’ is synonymous with ‘believes that’? First, note that they have the same extension; the extension of both expressions can be represented by the very same set of ordered pairs of subjects and propositions. (The expressions are true of the very same things.) Still, one might claim that despite sameness in extension, when ‘ $y$ ’ in ‘**T**( $x$ ,  $y$ )’ is replaced with a sentence, the resulting formula is not false but nonsensical. But this seems wrong. As long as the claim that someone believes a truth-value *makes sense* – as long as, say ‘God believes the truth’ *can* be interpreted as saying

(perhaps falsely) that God is related by belief to a certain non-intensional object (i.e., the truth) – we can coherently stipulate that both ‘believes that’ in ‘Bela believes that Opus 132 is a masterpiece’ and ‘T’ in ‘T (Bela, Opus 132 is a masterpiece)’ are different expressions with the same sense. ‘Believes that’ and ‘T’ both express {belief} and have the relation of belief as their extension.

We now consider two expressions: ‘Bela believes that Opus 132 is a masterpiece’ and the partially artificial ‘T (Bela, Opus 132 is a masterpiece)’. Again, according to Frege’s theory, ‘Opus 132 is a masterpiece’ denotes a truth-value. So, given the transparency of ‘T’, ‘T (Bela, Opus 132 is a masterpiece)’ expresses the proposition that Bela believes a particular truth-value.<sup>10</sup> Three claims allow us to draw the absurd result that the English sentence ‘Bela believes that Opus 132 is a masterpiece’ also expresses the proposition that Bela believes a particular truth-value.

- (a) We have assumed for our reductio that the sense expressed by ‘Opus 132 is a masterpiece’ does not differ between oblique and transparent contexts. So ‘Opus 132 is a masterpiece’ expresses the very same sense in ‘T (Bela, Opus 132 is a masterpiece)’ that it does when it occurs in ‘Bela believes that Opus 132 is a masterpiece’.
- (b) ‘Bela’ occurs unembedded in both ‘Bela believes that Opus 132 is a masterpiece’ and ‘T (Bela, Opus 132 is a masterpiece)’ so its sense does not differ between the two expressions.
- (c) We have argued above that there is nothing barring us from stipulating that ‘T’ and ‘believes that’ have the same sense.

We can therefore conclude that the proposition expressed by ‘T (Bela, Opus 132 is a masterpiece)’ is composed of the same constituents in the same manner as is the proposition expressed by ‘Bela believes that Opus 132 is a masterpiece’. If propositions composed of the same constituents in the same manner are identical to one another, ‘Bela believes that Opus 132 is a masterpiece’ and ‘T (Bela, Opus 132 is a masterpiece)’ express the same proposition. Given Frege’s extensional semantics, replacing ‘Opus 132 is a masterpiece’ as it occurs in a larger expression with any materially equivalent sentence will preserve the truth-value of that larger expression.<sup>11</sup> So the truth of ‘T (Bela, Opus 132 is a masterpiece)’, and thus the truth of ‘Bela believes that Opus 132 is a masterpiece’, entails that Bela believes every true proposition. But this is absurd. If ‘Opus 132 is a masterpiece’ denotes a sense when it is embedded in ‘Bela believes that Opus 132 is a masterpiece’ this sense must differ from the sense it there expresses.

The above argument is aimed at a Fregean view according to which ‘believes that’ expresses a function from ordered pairs of subjects and propositions to truth-values. We now consider an alternative view, the second of the two Fregean theories on offer. Suppose that ‘believes’ is both extensional and transparent, but when ‘that’ is applied to a sentence, the resulting clause denotes a proposition. On this view ‘that Opus 132 is a masterpiece’ has the same denotation in transparent contexts (as in ‘It is true that Opus 132 is a masterpiece’) as it does in oblique contexts (as in ‘Bela believes that Opus 132 is a masterpiece’). Moreover, according to the theory on hand, ‘that’ does not itself express a sense when it attaches to sentences. There is no mode of presentation associated with ‘that’ when it is used in this way; in such contexts ‘that’ does not express a concept of some individual, property, kind or extension. Instead, the semantic role or function of ‘that’ (when it is to attached to a sentence) is simply to shift the reference of that sentence’s constituent expressions to their customary senses. When attached to sentences, ‘that’ functions in the same way as does our ‘{ }’ notation.<sup>12</sup>

If we assume the truth of this Fregean theory, we must alter the above argument (though minimally) to show that Fregeans are committed to a hierarchy. We now consider an artificial language  $L^*$  that does not contain ‘that’ or any comparable word, but is otherwise identical to English. Importantly,  $L^*$  contains an expression ‘believes’ that is synonymous with the English word ‘believes’ and plays the exact same semantic role as does ‘believes’. In particular, ‘believes’ does not ‘take up’ the role that (we have assumed) is played by ‘that’ on its ‘sentential’ use in English. We unproblematically stipulate that ‘Bela’ and the  $L^*$  expression ‘Bela’ are synonymous, and so are the expressions ‘Opus 132’ and ‘Opus 132’, and ‘is a masterpiece’ and ‘is a masterpiece’. By compositionality ‘Opus 132 is a masterpiece’ and ‘Opus 132 is a masterpiece’ express the same proposition. By hypothesis there is no occurrence of ‘that’ or an equivalent expression – not even a “hidden” occurrence of such an expression – in the  $L^*$  sentence ‘Bela believes Opus 132 is a masterpiece’. If we assume with the non-hierarchical theorist that ‘that Opus 132 is a masterpiece’ expresses the same proposition in ‘Bela believes that Opus 132 is a masterpiece’ that it there denotes, the propositions expressed by ‘Bela believes that Opus 132 is a masterpiece’ and ‘Bela believes Opus 132 is a masterpiece’ are composed of the same senses in the same manner. By compositionality, the two sentences therefore express the same proposition. But the later sentence says that Bela believes the true. So



the assumption of a non-hierarchical view of senses again leads us into absurdity.

#### 4. A HIERARCHICAL REJECTION OF CONCEPTUAL SELF-VERIFICATION

It seems then that the Fregean is committed to a hierarchy of thought contents; but it is equally clear that hierarchical views are at least consistent with a rejection of the self-verifying nature of cogito judgments. Suppose the hierarchy exists and let  $\{\{\text{John}\}\}$  be one of several second-level MOPs determining the first-level MOP  $\{\text{John}\}$ . If judgment entails consideration and considering a proposition entails considering its constituents, the fact that *S* has judged  $\langle\{S\}, \{\text{thinking}\}, \langle\{\{\text{John}\}\}, \{\{\text{niceness}\}\}\rangle\rangle$  entails that she has considered  $\langle\{\{\text{John}\}\}, \{\{\text{niceness}\}\}\rangle$ , but given only classical logic it does not entail that she has considered  $\langle\{\text{John}\}, \{\text{niceness}\}\rangle$ .

Indeed, it would be intuitively false to suppose that thinking a second-level sense requires thinking the first-level sense it determines. Both  $\{\{\text{John}\}\}$  and  $\{\text{the only first-level subject-determining MOP denoted by an expression in this paragraph}\}$  determine  $\{\text{John}\}$ , but, intuitively, a subject can consider the second-level MOP  $\{\text{the only first-level subject-determining MOP denoted by an expression in this paragraph}\}$  without therein considering the first-level MOP  $\{\text{John}\}$ .<sup>13</sup> So *S* might be related by judgment to  $\langle\{S\}, \{\text{thinking}\}, \langle\{\text{the only first-level subject-determining MOP denoted by an expression in this paragraph}\}, \{\{\text{niceness}\}\}\rangle\rangle$  without being related by thought to  $\langle\{\text{John}\}, \{\text{niceness}\}\rangle$ .

#### 5. A FODORIAN ATTEMPT TO RECONCILE THE HIERARCHY WITH CONCEPTUAL SELF-VERIFICATION

Though the existence of the hierarchy appears to permit the possibility of false cogito judgments, one might hope to use substantive metaphysical claims about the nature of MOPs to dispel these appearances. To see how this might be done, suppose with Jerry Fodor (1998) that MOPs are types of Language of Thought (LOT) inscription. Whether Lois uses 'Clark Kent' or 'Superman' she always refers to the same man, but the MOP deployed when she uses 'Superman' is the LOT expression 'SUPERMAN', while the MOP deployed with 'Clark Kent' is 'CLARK KENT' in the LOT. Suppose too, in accordance with this psychologicistic account of senses, that '*R*

judges that *s* is true if and only if ‘*S*’ is written in *R*’s judgment box, where ‘*S*’ is the LOT sentence activated when *R* understands the declarative natural language sentence ‘*s*’. (Fodor allows that sometimes ‘*S*’ will be one of several MOPs with the content that *s*.)

Suppose Lois judges that she is thinking that Superman is heroic. What, according to the LOT theory of senses, is written in Lois’ judgment box? There are different proposals. According to the untenable non-hierarchical Fregean view (\*) is there written. According to the hierarchical view it is instead (\*\*) – or something similar – that appears there.

(\*) I AM THINKING THAT SUPERMAN IS HEROIC.

(\*\*) I HAVE ‘SUPERMAN IS HEROIC’ WRITTEN IN MY THOUGHT BOX.

Clearly, if Lois has ‘I AM THINKING THAT SUPERMAN IS HEROIC’ written in her judgment box, she must have ‘SUPERMAN IS HEROIC’ there written. And if judgment is a subspecies of thought – if the judgment box is a ‘compartment’ of the thought box – Lois must then have ‘SUPERMAN IS HEROIC’ written in her thought box. But suppose, instead, that ‘I HAVE ‘SUPERMAN IS HEROIC’ WRITTEN IN MY THOUGHT BOX’ is written in Lois’ judgment box, then, clearly, “SUPERMAN IS HEROIC” is there written, but because tokens of “SUPERMAN IS HEROIC” will literally contain tokens of ‘SUPERMAN IS HEROIC’, ‘SUPERMAN IS HEROIC’ will be written there as well. So even if there is a hierarchy of senses, if MOPs are linguistic entities of the Fodorian variety, cogito judgments will be self-verifying.

Unfortunately, however, on this reading, the LOT account would show too much, as it would entail that if *S* judges that she is thinking that *p*, *S* judges that *p*. If ‘I HAVE ‘SUPERMAN IS HEROIC’ WRITTEN IN MY THOUGHT BOX’ is written in Lois’ judgment box, then so is “SUPERMAN IS HEROIC” and ‘SUPERMAN IS HEROIC’. Thus, if we interpret the LOT theory and its box metaphor too literally, Lois cannot judge truly that she is *merely* entertaining the thought that Superman is heroic; and this is obviously untenable.

This is not to say that Fodor’s general theory of content is mistaken or that it cannot be augmented to account for the self-verifying status of cogito judgments. According to Fodor, the judgment or belief box is distinguished from, e.g., the desire box, by the differing functional roles of the LOT expressions each contains. So to save the

LOT account from absurdity, the Fodorian will surely say – the box metaphor notwithstanding – that when it is tokened in a subject's psychology 'I HAVE 'SUPERMAN IS HEROIC' WRITTEN IN MY THOUGHT BOX' might play the functional role of a judgment, without a token of 'SUPERMAN IS HEROIC' playing this role. Still, if the Fodorian is to explain conceptual self-verification she must go beyond the containment model and claim that it is a conceptual truth that whenever a token of 'I HAVE 'S' WRITTEN IN MY THOUGHT BOX' has the causal powers of a judgment, a fully distinct co-occurring token of 'S' has the causal powers of a thought. On the face of it, this fact about the respective causal powers of non-overlapping LOT expressions looks like a contingent truth, or (at best) a mere physical necessity. If its truth is contingent, then the Fodorian account entails that it is possible to judge that one is thinking that  $p$  without thinking that  $p$ . And if its truth is a mere physical necessity, the Fodorian account entails that it is conceivable that one judge that one is thinking that  $p$  without thinking that  $p$ . Moreover, the Fodorian faces arguments of a roughly Humean nature that the truth of the relevant claim *cannot* be conceptually necessary.<sup>14</sup> It is conceptually necessary that if an entity exists an essential part of that entity must exist as well. But if two concrete entities are mereologically distinct, isn't it at least conceivable that the one exist without the other? The Russellian can explain the conceptually self-verifying status of cogito judgments while respecting Hume's claim that there are no conceptually necessary connections between distinct concrete existences. The Fodorian cannot. Fodorians can have their Hume, and perhaps they can have their Descartes, but they cannot have both.

## 6. CHURCH'S VIEW

Still, the Fregean need not adopt a LOT account of MOPs. Alonzo Church (1951, 1975) provides a different way to reconcile the hierarchy with conceptual self-verification. Church's higher-level MOPs are conceived of as distinct from the lower-level MOPs they determine. (Or, at least, Church's formalism suggests that they are distinct.) Still, according to Church, (where  $n \geq 1$ ) each  $n$ th-level MOP is determined by a 'standard'  $(n + 1)$ th level sense, and it is a theorem of his logic – the Sense-Relationship Theorem (1974: 145) – that a standard higher-level MOP determines the particular lower-level

MOP for which it is standard. So let  $\{\{\text{John}\}\}$  be the standard sense that 'John' has in singly oblique contexts. Church claims that it is a truth of logic that  $\{\{\text{John}\}\}$  determines  $\{\text{John}\}$ .

Note, however, that even if true, Church's logical claim does not entail that someone who grasps  $\{\{\text{John}\}\}$  must (therein) grasp  $\{\text{John}\}$ , because 'grasping' is not closed under determination. Even if the fact that  $\{\{p\}\}$  determines  $\{p\}$  is entailed by logical theorems governing the determination relation, a subject who does not employ these theorems in a derivation (or apply them correctly) might grasp  $\{\{p\}\}$  without grasping  $\{p\}$ . Thus, to allow for the self-verifying status of cogito-judgments the Church-follower must swap premise (4) of our §1 argument with the following:

- (4') Where  $n \geq 1$ , to think a standard  $(n+1)$ th-level sense one must therein think the  $n$ th-level sense it determines. (For example, where  $\{\{p\}\}$  is the standard second-level sense determining  $\{p\}$ , for S to think  $\{\{p\}\}$  she must think  $\{p\}$ .)

Suppose we assume (4'). Then if  $S$  is related by thought to  $\langle\{S\}, \{\text{thinking}\}, \{\{\{\text{John}\}\}, \{\{\text{niceness}\}\}\rangle\rangle$ ,  $S$  must therein grasp and so be related by thought to  $\langle\{\text{John}\}, \{\text{niceness}\}\rangle$ . Moreover, (4') allows that  $S$  might be related by thought to  $\langle\{S\}, \{\text{thinking}\}, \langle\{\text{the only first-level subject-determining MOP denoted in this paragraph}\}, \{\{\text{niceness}\}\}\rangle\rangle$  without grasping  $\langle\{\text{John}\}, \{\text{niceness}\}\rangle\rangle$ .

Should we accept (4')? It is hard to say, if only because it is far from clear what Church's standard higher-level senses are supposed to be.<sup>15</sup> We can follow Church and denote a standard second-level sense with well-defined artificial notation, or we can follow Herbert Heidelberger (1975) and use quasi-natural expressions such as 'the sense of 'the sense of 'Opus 132' '' for this purpose. But our ability to denote higher-level senses in either of these ways is (in Russell's words) 'merely linguistic through the phrase': it does not provide us with substantive information about standard higher-level senses that we can use to evaluate claims about their cognitive role.

Thus, the Fregean must develop a special theory of 'standard' higher-level content if she is to account for the self-verifying status of cogito judgments. This special theory will have to say that, necessarily, a subject grasps a standard higher-level sense *by* grasping the lower-level sense it determines along with a concept or function from lower-level senses to their standard determiners. (For example, it might say that to grasp  $\{\{\text{John}\}\}$  one must grasp the function expressed by 'the standard sense determining  $x$ ' and apply this function

to {John}.) But the theory will have to say more than just this. Intuitively, a subject must grasp {animal} if she is to grasp {cat}, but this does not entail that when a subject entertains the proposition expressed by 'Whiskers is a cat', she therein entertains {animal}. {Cat} is a first-level sense that cannot be grasped without a great deal of conceptual development – it is probably the case that in the course of this development {animal} must be grasped. But {cat} is nevertheless plausibly thought of as *atomic*: it does not contain internal inference-underwriting structure. So, for example, the inference from 'Whiskers is a cat' to 'Whiskers is an animal' is *indirect*, unlike the inference from 'Whiskers is an animal with features  $F_1$ – $F_n$ ' to 'Whiskers is an animal'.<sup>16</sup> If this is right, the neo-Fregean cannot simply say that we come to grasp higher-level senses by grasping the senses they determine; she must say that higher-level senses are composite in nature and that (because of this composite structure) one cannot entertain a higher-level sense without therein – i.e., in the *same* act of mind – entertaining the lower-level sense(s) it determines. These theses constitute a substantial augmentation of the neo-Fregean theories currently on offer.

It should be clear, however, that the Russellian need not adopt a 'special fix' of this sort because the Russellian's general theory of content already entails that the contents of second-order judgments contain the relevant first-order contents within them. Thus, given plausible assumptions about the nature of judgment, thought and cognitive consideration – i.e., the assumptions made explicit in premises (1)–(3) – the Russellian theory of content entails that cogito judgments are self-verifying. The Fregean's explanatory position is drastically different. Because the Fregean must accept a hierarchy of senses, her theory is incompatible with (4), and in the absence of augmentation the Fregean theory does not entail (4'). Indeed, unless she explicitly adds Church's theorem and posits standard higher-level senses with the requisite internal structure there is nothing about the Fregean's theory of content that even argues in favor of (4').

## 7. CONCLUSION

It seems then that the conceptually self-verifying nature of cogito judgments places serious constraints on our theory of thought-content. Indeed, the tension goes beyond the writings of those who mistakenly endorse a non-hierarchical view of MOPs. The 'containment model' of the relation between first- and second-order

content has become a standard reply to arguments for the incompatibility of content externalism and privileged introspective access,<sup>17</sup> while Frege-inspired accounts of thought-content remain extremely popular. Insofar, as the arguments presented here are cogent, these two positions are difficult to reconcile. To retain the Cartesian (conceptual) explanation for why a thinker cannot judge falsely that she is thinking a particular thought, the Fregean must claim that only some of the (potentially infinite) higher-level MOPs that determine a single lower-level MOP are standard. And she must move beyond current formulations of Fregeanism to explain why the fact that one is grasping a standard higher-level MOP entails that one is concurrently grasping the lower-level MOP for which it is standard.

Of course, the Fregean who rejects (4') might nevertheless posit an intimate (perhaps causal) relation between the content of the thought that *p* and the content of the introspective judgment that one is thinking that *p*. But if neither (4) nor (4') is correct, cogito judgments are not conceptually self-verifying.

#### NOTES

<sup>1</sup> I would like to thank Anthony Brueckner, Eric Gilbertson, Carl Ginet, Delia Graff, Brendan Jackson, Keith McPartland, Christopher Peacocke, Michael Rescorla, Sydney Shoemaker and Jason Stanley for discussion of these and related issues, and the members of the 2002 NEH summer seminar on Consciousness and Intentionality where I presented a fairly distant ancestor of this paper. Special thanks go to C. Anthony Anderson, Kevin Falvey and Nathan Salmon for key references and particularly helpful discussions at several meetings of the Santa Barbarians, and to two referees from *Synthese* for valuable written comments.

<sup>2</sup> Every judgment with a necessarily true content is therefore trivially self-verifying. Strictly speaking, then, Descartes draws our attention to an interesting sub-category of self-verifying judgments – judgments that are made true by their occurrence: *S* judges that *p*, *p* is not a necessary truth, but the fact that *S* judges that *p* entails *p*. The class of judgments Burge discusses is even narrower than this.

<sup>3</sup> The argument generalizes somewhat. If *S* judges that she was thinking that *p*, then *S* is thinking that *p*. If *S* judges that she will be thinking that *p*, then *S* is thinking that *p*. The former judgment is not self-verifying because the time at which *S* judges that she was thinking that *p* may be the first time at which she considers the proposition that *p*; the latter judgment is not self-verifying because the time at which *S* judges that she will be thinking that *p* may be the last time at which she considers the proposition that *p*. Moreover, it is plausible to suppose that (like judging that *p*) doubting that *p* entails considering the proposition that *p*. If so, doubting that one is thinking that *p* entails thinking that *p*, and any such doubt is (as Descartes says) self-defeating. If one knows that doubting entails thinking, one can infer that the very

proposition that one doubts when one doubts that one is thinking that  $p$  is in fact a true proposition (so long as one is doubting it). A rational agent who drew this inference would be forced to abandon her doubt.

<sup>4</sup> What about ‘unnatural’ combinations of constituents? Does  $S$ ’s thinking that John is nice and Mary is more fashionable than Bill entail  $S$ ’s considering (fashionable, nice, Bill)? If not, we can weaken (2) to say that considering a proposition entails considering its truth-evaluable (or judgable) constituents. But does  $S$ ’s thinking what she does entail  $S$ ’s considering (More fashionable than (Bill, John))/? If not, we can either: (a) claim that (More fashionable than (Bill, John)) is not a constituent of the proposition that John is nice and Mary is more fashionable than Bill, or (b) weaken (2) still further to say that considering a content entails considering its ‘order-preserving’ judgable constituents. The above argument for the self-verifying status of cogito judgments goes through on either revision.

<sup>5</sup> I here make the (surely false) simplifying assumption that ‘John’ has a single customary MOP. This will not affect what follows.

<sup>6</sup> I here define hierarchical views as those that accept at least two levels of senses. (On my terminology, theorists who have tried to limit the hierarchy to exactly two levels have nevertheless admitted a hierarchy.) Peacocke is not alone in adopting a non-hierarchical version of Frege’s theory of content. As he points out, Carnap (1947, p. 129) and Dummett (1973, pp. 267–269, 1981, pp. 83–147) endorse the view as well.

<sup>7</sup> The argument below owes a great deal to Nathan Salmon’s discussion in ‘On Designating’, of the famous Gray’s ‘Elegy . . .’ passage in Russell’s ‘On Denoting’. Compare with Blackburn and Code (1978). I should also make it clear that the argument that follows is not a general refutation of the redeployment claim. Instead, it shows that the redeployment claim is incompatible with a Fregean account of content.

<sup>8</sup> The non-hierarchical Fregean might object to this assumption. (Peacocke would not, as he employs the same ‘{ }’ notation to denote senses that we employ above.) But, first, it is hard to see why we should believe in the existence of entities we cannot simply denote; and, second, this view is not available to neo-Fregeans – like Burge – who wish to make substantive claims about particular concepts (e.g., {arthritis}). Cf. Church, ‘The hierarchy of concepts of successively higher orders arises as soon as we suppose that a concept, like anything else that can be discussed at all, is capable of having a name given to it’ (1951, p. 12, n. 13).

<sup>9</sup> This must be Dummett’s view as he admits that if embedded expressions express the very senses they denote, the sense of ‘Aristotle’ in *oratio obliqua* cannot also be equated with the ordinary sense of the phrase ‘the sense of ‘Aristotle’’ (1981: 89). If there are senses, ‘the sense of ‘Aristotle’’ clearly has a sense; if this sense is not the sense of ‘Aristotle’ it must be a second-level sense that determines the sense of ‘Aristotle’. So Dummett must allow that there is a hierarchy, even if only its first-level members enter into the contents of higher-order thoughts.

<sup>10</sup> Burge says, “‘Believes’ originally applied to person and propositions. But on our assumptions it has come to apply to persons and truth-values’ (1979, p. 272).

<sup>11</sup> For the claim that the truth-value of a sentence is a function of the denotations of its parts see Frege (1980, p. 59).

<sup>12</sup> We here make no assumptions about the semantic value of ‘that’ when it is attached to sub-sentential expressions and so functions as a demonstrative. To be fair, one might think that even when ‘that’ is attached to a sentence it functions as a

demonstrative. That is, one might think that in such contexts, 'that' expresses a demonstrative sense determining the proposition expressed by the sentence that follows it. I don't think this view is plausible, but it can be ignored in the present context because it requires positing second-level senses in oblique contexts (i.e., demonstrative MOPs that are distinct from the propositional MOPs they determine) and so cannot be adopted by a non-hierarchical Fregean. There is also Davidson's (1968–1969) view that tokens of 'that' demonstrate (or are used to demonstrate) the token utterances they precede. Davidson's view is unacceptable to hierarchical and non-hierarchical Fregeans alike.

<sup>13</sup> {John} is one of the potentially infinite MOPs that determine John and it is the customary MOP expressed by the English expression 'John'. {The only first-level subject-determining MOP denoted by an expression in this paragraph} is one of the potentially infinite MOPs that determine the only first-level subject-determining MOP denoted by an expression in this paragraph – i.e., {John} – and it is the customary MOP expressed by the somewhat technical English expression 'the only first-level subject-determining MOP denoted by an expression in this paragraph'.

<sup>14</sup> See Armstrong (1963).

<sup>15</sup> Cf. Carnap, 'since [Frege] assumes that nominatum and sense must always be different, he had thus to introduce a third entity as the oblique sense. Incidentally, it seems that Frege nowhere explains in more customary terms what this third entity is' (1949, p. 129).

<sup>16</sup> See Fodor (1998) and Fodor et al. (1999).

<sup>17</sup> Among the prominent advocates of the containment model are Heil (1988) and Shoemaker (1996).

## REFERENCES

- Armstrong, D.: 1963 'Is Introspective Knowledge Incorrigible?'; *The Philosophical Review* **72** (4), 417–432.
- Blackburn, S. and A. Code: 1978, 'The Power of Russell's Criticism of Frege: 'On Denoting,' pp. 48–50,' *Analysis* **38**, 65–77.
- Burge, T.: 1977 'Belief *De Re*,' *Journal of Philosophy* **74**, (6), 338–362.
- Burge, T.: 1979 'Frege and the Hierarchy,' *Synthese*, **40**, pp. 265–81.
- Burge, T.: 1988 'Individualism and Self-Knowledge,' *Journal of Philosophy* **85**, (11), 649–663.
- Carnap, R.: 1947 *Meaning and Necessity*, UP, Chicago.
- Church, A.: 1951 'A Formulation of the Logic of Sense and Denotation,' in P. Henle, H. Kellen and S. Langer (eds.), *Structure, Method And Meaning: Essays in Honor of Henry M. Sheffer*, Liberal Arts Press, New York, pp. 3–24.
- Church, A.: 1974 'Outline of a Revised Formulation of the Logic of Sense and Denotation (Part II),' *Nous* **8**, 135–156.
- Davidson, D.: 1968–9 'On Saying That,' *Synthese* **19**, 158–174.
- Dummett, M.: 1973 *Frege: Philosophy of Language*, Duckworth, London.
- Dummett, M.: 1981 *The Interpretation of Frege's Philosophy*, Harvard UP, Cambridge, MA.
- Fodor, J.: 1998 *Concepts: Where Cognitive Science Went Wrong*, UP, Oxford.



- Fodor, J. A., M. Garrett, E. Walker, and C. Parkes: 1999 'Against Definitions,' in E. Margolis and S. Laurence (eds.), *Concepts: Core Readings*, MIT Press, Cambridge, MA, pp. 491–512.
- Frege, G.: 1980; first published 1892 'On *Sinn* and *Bedeutung*,' in P. Geach and M. Black (eds.), M. Black (trans), 3<sup>rd</sup> edn., *Translations From the Philosophical Writings of Gottlob Frege*, Blackwell, Oxford, pp. 56–78.
- Heil, J.: 1988 'Privileged Access,' *Mind* **97**, 238–251.
- Heidelberger, H.: 1975 'Review of Frege: *Philosophy of Language*,' *Metaphilosophy* **6**, 35–43.
- Peacocke, C.: 1999 *Being Known*, Clarendon, Oxford.
- Salmon, N. 'On Designating,' forthcoming in *Mind*.
- Shoemaker, S.: 1996 *The First-Person Perspective and Other Essays*, UP, Cambridge.

University of California  
Department of Philosophy  
5707 South Hall  
Santa Barbara, CA 93106  
USA