

# SAYING NOTHING AND THINKING NOTHING

John A. Keller and Lorraine Juliano Keller  
Saint Joseph's University

## 1. INTRODUCTION

Lapsing into nonsense is an occupational hazard of philosophy. But, unless they've been drinking, the sort of nonsense philosophers are liable to lapse into is (usually) not pure gibberish—rather, it's nonsense that has the illusion of making sense (“deceptive nonsense”). Deceptive nonsense is sometimes accompanied by what Gareth Evans (1982) called “illusions of thought”: cognitive events that *seem* to have content, but don't. But if nonsense sentences, assertions, and thoughts don't mean *anything*, it's hard to see how such illusions could arise. As Carnap famously asked,

And how could one account for the fact that metaphysical books have exerted such a strong influence on readers up to the present day, if they contained not even errors, but nothing at all? (1959: 78)

In this paper we defend the existence of deceptive nonsense and illusions of thought by (i) sketching a general framework for thinking about them, (ii) clarifying the sense in which they lack meaning, (iii) providing arguments for their existence, and (iv) responding to some arguments against them.

## 2. OVERVIEW

In her classic *Encyclopedia of Philosophy* article, Annette Baier distinguishes six species of nonsense, claiming that an utterance is nonsense if it has one or more of the following features:<sup>1</sup>

1. It is obviously false
2. It is wildly inapposite
3. It involves a category error
4. It is syntactically ill-formed
5. It is an otherwise meaningful sentence containing nonsense words (meaningless words)
6. It is a string of nonsense words.

Examples of each species include:

1. The dog is a mathematician.
2. The dog chased the cat [spoken by a defendant in court, in answer to the question of where they were at the time of the crime].
3. The piano has been drinking.
4. cats Blargs chase.
5. Blargs chase cats.

---

<sup>1</sup> See Baier (1972: 521). She calls these the “main” ways of departing from sense, leaving open whether there are other ways. We avoid the word ‘sentence’ here because it's controversial that all of the relevant examples count as sentences.

6. Slithy toves brillig.

One way of grouping these species into genera would be to contrast what Keller (2017) calls **silly nonsense**—utterances whose contents are bizarrely false, like ‘I have 2.3 children’, or perhaps ‘Caesar is a prime number’—with what he calls **semantic nonsense**: utterances that fail to have a meaning and hence cannot be evaluated for truth or falsity at all. (It is a matter of some dispute whether category mistakes—nonsense belonging to species 3 in Baier’s taxonomy—belong to the genus of silly or semantic nonsense.<sup>2</sup>) Examples (4)-(6) are semantic nonsense, since (5) and (6) contain meaningless expressions and (4) combines words in a meaningless way. Examples (1) and (3), on the other hand, are silly nonsense. It would be nice if we could fit (2) into the category of silly nonsense, but it probably belongs to a genus of its own. Our focus is on semantic nonsense: specifically, examples like (4)-(6).<sup>3</sup> Or rather, we’re interested in examples that are like (4)-(6) in lacking meaning, but unlike (4)-(6) in not *obviously* lacking meaning. Utterances like (4)-(6) are what might be called **gibberish**: semantic nonsense that is *obviously* nonsense; nonsense that no vaguely competent speaker could think was anything other than nonsense. Consider, however,

7. Das Nichts nichtet.<sup>4</sup>

8. Vulcan is a planet.<sup>5</sup>

9. Witches cast spells.<sup>6</sup>

These are sentences that some people have sincerely uttered but that others think lack meaning. If these sentences are nonsense, they are (at least potentially) **deceptive nonsense**. While controversial assumptions are required to deliver the verdict that (7), (8), and (9) are nonsense, examples of deceptive nonsense can be generated from less contested principles. Consider Maeve (age 4), and Ava and Jake (her parents), in a context where nobody named ‘John’ is salient.

**Case 1:** Suppose Ava and Jake are talking about names for their new baby, and Ava suggests ‘John’. In response, Jake says, “‘John’ is abominable.” Maeve, passing by, hears Jake but doesn’t realize that the name ‘John’ is being mentioned rather than used. Excited to learn this bit of apparently salacious news, she rushes off and tells her friend Izzy (who believes her), uttering:

10. John is abominable.

---

<sup>2</sup> See Magidor (2013) for discussion.

<sup>3</sup> Cora Diamond (1981: 10) has argued that according to (certain) forms of Fregeanism, category (5) collapses into category (6). Her thought is that if the Context Principle is true—if words only have meanings when they are embedded in (meaningful) sentences, since the meanings of words are partially *determined* by the meanings of the sentences in which they appear—then *none* of the words in ‘Blargs chase cats’ means anything, since the sentence itself doesn’t mean anything.

<sup>4</sup> See Carnap 1959, discussing Heidegger (1929). Often translated as ‘the nothing noths’, this is widely viewed as a neologism that makes no grammatical or conceptual sense.

<sup>5</sup> See Braun 1993. Since ‘Vulcan’ is an empty name, it has no content on Millian views where the contents of names are their referents. This plausibly entails that the sentence as a whole lacks content. (But see fn.11 below.)

<sup>6</sup> See Braun 2015. Plausibly, ‘witch’, ‘cast’, and ‘spell’ lack determinate meanings: since there are no witches, acts of casting, or spells, we cannot rely on the world to settle ambiguities and contradictions in how these terms are used and in the concepts we associate with them.

**Case 2:** Discussing potential names, Jake suggests ‘Donald’, and Ava says in reply, “Your suggestion is abominable.” Maeve, passing by halfway through Ava’s utterance, thinks she hears Ava say (10), and rushes off to tell her friend Izzy, who believes her.

**Case 3:** Maeve (age 8 now) comes across a piece of paper with some derogatory statements printed on it, including (10), and takes those sentences to be assertions by some speaker about somebody named ‘John’. Maeve then repeats (10) to Izzy, intending to defer to that speaker about the referent of ‘John’. Sadly, however, those inscriptions were randomly generated by a computer, and there is nobody that ‘John’ refers to.

The recipe for generating cases of deceptive nonsense like (1)-(3) is simple: a speaker  $S$  (thinks she) learns a name  $N$  from some other speaker  $S^*$  (by deferring to  $S^*$ ’s referential intentions), but unbeknownst to  $S$ ,  $S^*$

- a) was mentioning  $N$  (the word, without an intended referent)
- b) was using another expression  $N^*$  (in another grammatical category) that  $S$  mistakes for  $N$
- c) does not exist ( $N$  isn’t actually produced by a speaker)
- d) was engaging in deliberate nonsense (had no intended referent)

In such cases, by attempting to defer to  $S^*$ ,  $S$  ensures that her use of the name is meaningless, and that the sentences uttered by  $S$  using that name will fail to express propositions.<sup>7</sup>

One interesting thing about deceptive nonsense is that it can give rise to:

**Illusions of Meaning:** sentences that seem to have meanings, but that really don’t.

**Illusions of Assertion:** speech acts that seem to say or assert something, but that really don’t.

**Illusions of Thought:** cognitive episodes that seem to have contents, but that really don’t.

By definition, a sentence  $s$  that is deceptive nonsense has the illusion of being meaningful; and so an utterance of  $s$  would, in typical cases, create an illusion of assertion, and giving one’s assent to such an utterance—or denying it, for that matter—would, it seems, typically result in an illusion of thought. (See §2.2 for a discussion of the need for hedging these claims.) Furthermore, as Jim Pryor writes, ‘Rehearsing sentences to yourself is one way of having occurrent thoughts’ (2006: 329 n. 1) Since such silent talking to ourselves is a cognitive episode, it follows that if one silently rehearses  $s$  to oneself (while taking  $s$  to be meaningful), one is suffering from an illusion of thought.

**2.1 Kinds of Meaning.** We’ve said that sentences (and assertions and thoughts) that fail, in whole or in part, to have meanings are semantic nonsense. But ‘meaning’ is said in many ways. The relevant kind of meaning that nonsense expressions lack is *content*.

---

<sup>7</sup> *Mutatis mutandis* for nonsense predicates. Consider **Case 4:** Ava says to Jake, “What’s an example of nonsense?”, and Jake says in reply

11. Geeshjohn is minabobable.

Maeve, passing by, hears Jake’s utterance of (11), which she then repeats to Izzy, who believes her.

There is wide disagreement about the *nature* of content, but wide agreement about the roles it plays. Contents that are *truth-apt*—susceptible of truth and falsity—are commonly referred to as **propositions**. Propositions are the fundamental bearers of truth value, the semantic contents of univocal, declarative sentences (in contexts of utterance), and the contents of thoughts (beliefs, hopes, etc.) and assertions. Sentences are true (or false) by virtue of expressing true (or false) propositions. Sub-sentential expressions—lexical items, such as ‘girl’ and phrases, such as ‘down the street’—do not have propositions as contents. The specific nature of sub-sentential content is contentious (see §3.1-§3.3), but the contents of sub-sentential expressions are generally taken to determine, in conjunction with syntax, the propositions expressed by sentences in which those expressions occur. Typically, if a constituent of a sentence lacks content, the sentence itself will fail to express a proposition, as is the case in (5) above: since ‘blarg’ does not have a content, (5) does not express a proposition. (If an argument for this claim is needed, note that (5) cannot be evaluated for truth or falsity.)

Because natural languages like English contain indexicals (‘I’, ‘now’) and demonstratives (‘that’, ‘those’) the semantic contents of which are partially determined by the context in which they are used, content must be distinguished from **standing meaning**.<sup>8</sup> Standing meaning is the context-invariant linguistic information associated with *expression-types*—the conventional, linguistic rules mastery of which characterizes language acquisition. But the standing meaning of expression-types is not always sufficient for determining content.

For example, the sentence type ‘I am hungry’ does not express a proposition (a truth-evaluable content), since it might be true when Maeve utters it and false when Izzy does (or true when Maeve utters it before lunch and false when she utters it after): that is, some tokens of it are true and others are false. Whether it is true or false depends on features of the context, such as the speaker and time of the utterance. Content, then, is the sort of meaning that is associated with sentence *tokens* in contexts of utterance. In general, the semantic content of an expression is determined by its standing meaning, plus the relevant features of context (e.g., speaker, time, and place). In what follows, we use double brackets as a device for referring to the content of the expression within: just as putting quotation marks around an expression *e* creates a name for *e*, putting double brackets around a denoting expression *e* creates a name for the content of *e*’s denotation. Hence, e.g., [[‘Grass is green’]] = the proposition that grass is green (that is, [‘Grass is green’] refers to the proposition that grass is green).

One hallmark of semantic nonsense is that it is not truth-evaluable, even in context. That’s why the most important sense in which semantic nonsense lacks meaning is that it lacks content. Of course, any sentence that lacks a standing meaning will also lack content, since content is determined by standing meaning (in context). Certain paradigm instances of nonsense—what we have called “gibberish”—lack a standing meaning, and hence lack content. Because gibberish lacks a standing meaning, it is generally not deceptive nonsense: since gibberish lacks *any* sort of linguistic meaning, speakers (and listeners) will generally *know* that gibberish lacks content. But it is not always the case

---

<sup>8</sup> See Heck (2002).

that (even relatively competent) speakers know whether expressions lack a standing meaning (according to many views about standing meaning): ‘John’ (in (10)) is not gibberish to Maeve, for example, even though it lacks both a standing meaning and a content. So we cannot simply say that the difference between gibberish and deceptive nonsense is that gibberish lacks a standing meaning in addition to lacking content.

**2.2 Kinds of Semantic Nonsense.** Propositions, as fundamental bearers of truth-conditions, are the (potentially) shared contents of sentences, assertions, and thoughts. For example, the sentence, ‘Ava is human’, Izzy’s assertion of that sentence, and Izzy’s belief that Ava is human all have the proposition [[‘Ava is human’]] as their content. This partly explains why we use that sentence to report Izzy’s belief. Since the bearers of truth-evaluable content (things that express propositions) fall into three main categories, there are three main categories of nonsense:<sup>9</sup>

**Sentential Nonsense:** a *sentence* that lacks content.

**Assertoric Nonsense:** an *assertion* that lacks content.

**Cognitive Nonsense:** a *thought* (belief, hope, desire, etc.) that lacks content.

Note that these three types of semantic nonsense can give rise to the three types of illusion outlined in §2: sentential nonsense can give rise to illusions of meaning, assertoric nonsense to illusions of assertion, and cognitive nonsense to illusions of thought. It is worth noting, however, that these different kinds of nonsense do not necessarily go hand-in-hand. Depending on one’s views about about assertion, it might be possible to have sentential nonsense without assertoric or cognitive nonsense, since we don’t put our thoughts into words perfectly: one could, e.g., misspeak in a way that results in sentential nonsense, perhaps by speaking ungrammatically. Still, there would be a thought one was trying to express, and one’s interlocutors might readily grasp one’s meaning (and so, on at least some theories of assertoric content, one would have performed a contentful speech act). For example, the second sentence before this one is ungrammatical, and hence (arguably) does not express a proposition, even in context. But there was a thought we were trying to communicate when we wrote it, and we plausibly managed to successfully assert or express that thought. After all, you probably took our meaning: indeed, you probably didn’t even notice the error.<sup>10</sup> If small grammatical errors prevented us from saying or asserting anything, we could gain the benefits of lying without actually lying (defined as asserting something we correctly believe to be false) by inserting slight grammatical infelicities into our speech.

Conversely, if (mental) content externalism is false but semantic externalism is true, there will be cases where people sincerely utter meaningful sentences but only have the illusion of thought, since they don’t really know or understand the meaning of the sentences they’re uttering. When students first learn about Einstein’s theory of relativity, they learn to say things like, ‘Simultaneity is relative to reference frame’. That sentence is meaningful and true, as are assertions of it. But it’s unlikely that

---

<sup>9</sup> For related discussion, see Cappelen (2013: 26).

<sup>10</sup> If you don’t see it, ‘about’ is repeated.

beginning students actually know what the sentence means: there's no thought or belief of theirs that that sentence expresses. They are merely parroting their teachers.

Such students may suffer from what Keller (2017: 2) calls "illusions of nonsense": cases where something meaningful or true seems to be semantic or silly nonsense (perhaps a category mistake). As an example of an illusion of *silly* nonsense, consider, 'There are as many even numbers as numbers', a sentence that a precocious 7<sup>th</sup> grader might dismiss out of hand. Examples of illusion of *semantic* nonsense can be brought to mind by thinking of cases where someone conflates the fact that *she* doesn't know what an expression means with the expression's lacking a meaning. This is...not uncommon in philosophy.

### 3. IN FAVOR OF DECEPTIVE NONSENSE AND ILLUSIONS OF THOUGHT

It's trivial to provide examples of gibberish: we already have. But those examples weren't assertions; they were *deliberate* nonsense. The interesting questions are (i) whether there is *deceptive* nonsense: utterances that are intended to be meaningful, and that one takes to be meaningful, but which nonetheless fail to be meaningful; and (ii) whether such deceptive nonsense is accompanied by illusions of thought. The cases we gave in §2 suggest so, as does the testimony of philosophers. Consider:

- ❑ "Natural rights is simple nonsense: natural and imprescriptible rights, rhetorical nonsense—nonsense upon stilts." (Bentham 1843)
- ❑ "The book will...draw a limit to...the expression of thoughts....The limit can...only be drawn in language and what lies on the other side of the limit will be simply nonsense" (Wittgenstein 1961: Preface)
- ❑ "The alleged statements of metaphysics which contain [words like 'God' and 'essence'] have no sense, assert nothing, are mere pseudo-statements" (Carnap 1959: 67)
- ❑ "Philosophical problems arise when language goes on holiday." (Wittgenstein 1953: §38)
- ❑ "It is obviously a perfectly significant statement, whether true or false to say that Romulus existed. If Romulus himself entered into our statement, it would be plain that the statement that he did not exist would be nonsense, because you cannot have a constituent of a proposition which is nothing at all." (Russell 1956: 242)
- ❑ "If a sentence makes no statement at all, there is obviously no sense in asking whether what it says is true or false. And we have seen that sentences which simply express moral judgments do not say anything. They are pure expressions of feeling and as such do not come under the category of truth and falsehood." (Ayer 1952: 108).
- ❑ "Reflective persons unswayed by wishful thinking can themselves now and again have cause to wonder what, if anything, they are talking about." (Quine 1960: 242)
- ❑ "I do not understand what philosophers say...and I think the reason I do not understand them is that they have failed to explain what they mean...And I think the reason they have failed to

explain what they mean is that there is nothing, or nothing coherent, that they do mean...” (van Inwagen 1980: 285)

- “Lacanian believe that the unconscious is structured like a language. They are not sure what this means, but they trust Lacan, who said so.” (Recanati 1997: 84)
- “[I have made] the *prima facie* case that uses of ‘intuitive’ and cognate terms are so defective that they should be classified as nonsensical.” (Cappelen 2013: 40)

This is but a small sample of the accusations of nonsense philosophers have leveled against each other. While we don’t agree with all of them, we do think that deceptive nonsense is not only possible, but actual, and probably common. That is, we think that philosophers (and others) have unintentionally spoken nonsense and suffered from the illusions that typically accompany it. Even if the philosophically interesting accusations of nonsense listed above are all incorrect, the mundane cases given in §2 seem to illustrate how easy it is for deceptive nonsense to arise. In the remainder of this section we support this intuitive verdict by showing how it can be vindicated from a number of different theoretical perspectives.

**3.1 Millianism.** According to Millianism, the semantic content of a name is its bearer. Thus, Millianism is a version of Direct Reference semantics, according to which the sole semantic function of a certain class of expressions (proper names, and perhaps indexicals and demonstratives) is to refer to an individual. Millianism entails that empty names—names without bearers, such as ‘John’ in (10), or ‘Vulcan’—lack semantic content. Since the semantic content of a sentence is determined by the semantic contents of its constituents (plus syntax), sentences containing empty names will plausibly lack content as well. As David Braun says,

“According to Direct Reference, if ‘Vulcan’ does not refer, it has no semantic value. Even worse, it seems that sentences containing ‘Vulcan’ cannot express propositions, since there is no semantic value to “fit into the subject position” of the proposition” (1993: 451).

To get from this view of empty names to deceptive nonsense and illusions of thought, all that is needed is the anodyne observation that sometimes we don’t know that a name is empty. Thus, Millianism leads, fairly straightforwardly, to deceptive nonsense and illusions of thought.<sup>11</sup>

**3.2 Neo-Fregeanism.** The existence of deceptive nonsense isn’t tied to Millian semantics, however. Any view about the contents of singular terms according to which they are *object-dependent* will entail that content is lacking in these cases. According to Neo-Fregean theories, for example, the contents of singular terms are the senses of those terms rather than their referents, but those senses are object-

---

<sup>11</sup> It is worth noting, however, that some Millians (including Braun 1993) defend a view according to which sentences like ‘Vulcan is a planet’ express “gappy propositions”. On such views, ‘Vulcan is a planet’ isn’t semantic nonsense as we’ve defined it, since it has content. There’s still a cognitive and semantic illusion though, since ‘Vulcan is a planet’ doesn’t have the kind of content that it seems to have. Its content is a gappy proposition, rather than a fully-saturated one, and it’s the *same* gappy proposition as the one expressed by ‘Phlogiston is a planet’. On this view, then, someone might still suffer from the illusion of thinking that [[‘Vulcan is a planet’]] and [[‘Phlogiston is a planet’]] were different.

dependent.<sup>12</sup> On such views, ‘John’ (in (10)) and ‘Vulcan’ will lack senses as well as referents. Thus, deceptive nonsense and illusions of thought arise on Neo-Fregean theories.

**3.3 Fregeanism.** On traditional Fregean views, senses are not object-dependent, and so empty names will generally have contents, since they will have senses. But only generally: on Fregean views, it’s still *possible* for a speaker to unknowingly use a name without a sense and thus fail to express a proposition. Cases (1)-(3) illustrate this. All that is required for the existence of deceptive nonsense on Fregean views is for *semantic deference* to be possible: for it to be possible to use an expression to mean whatever other speakers use it to mean (if anything). Failed attempts at deference will thus lead to a lack of content. In Cases (1)-(3), nobody was actually using ‘John’, and no sense of ‘John’ was expressed or made contextually salient; hence there was no content for Maeve’s use of ‘John’ to inherit, her deferential intentions notwithstanding. Thus, deceptive nonsense and illusions of thought arise on Fregean theories.<sup>13</sup>

**3.4 The Big Picture.** So far, we’ve considered cases involving expressions having too few, i.e., zero, meanings. So long as our attempts to introduce meaning can go wrong—and what in human affairs can’t go wrong?—there will be failed introductions, resulting in expressions with too few (i.e., zero) meanings. And so long as things can go wrong in this way without our *immediately* realizing it—and what can’t go wrong without it being immediately realized?—we will get cases of deceptive nonsense. As Cappelen (2013: 32) puts it, no plausible semantic or meta-semantic theory has a “built-in *guarantee of infallibility*” with respect to the introduction of new meaningful expressions, or our beliefs about whether such introductions have been successful.

But deceptive nonsense can arise from the other direction as well: not from an expression having too few meanings, but too many. If Maeve defers to Ava and Jake about the referent of ‘John’, but Ava and Jake are (perhaps unwittingly) using ‘John’ to refer to different people, Maeve’s deference will result in her use of ‘John’ lacking (a determinate) content. More generally, if Maeve intends to use an expression to mean whatever others are using it to mean, but there are multiple things others have used it to mean and neither Maeve nor her circumstances have done anything to differentiate between those possible meanings, her use will lack a determinate content. This is, roughly, Cappelen’s view of the word ‘intuition’ as used by many philosophers.<sup>14</sup>

#### 4. AGAINST DECEPTIVE NONSENSE AND ILLUSIONS OF THOUGHT

---

<sup>12</sup> This view is defended by, e.g., Gareth Evans (1982) and John McDowell (1986), perhaps the most (in)famous defenders of illusions of thought.

<sup>13</sup> We do not claim that this assessment is true to the historical Frege—we are not engaged in exegesis. Most likely, our analysis conflicts with Frege’s more stringent requirements on semantic competence (see Frege 1956), which are not widely held today. But even if Maeve isn’t competent with ‘John’, that wouldn’t affect the plausibility of the claim that Maeve thinks she is thinking and talking about (somebody named) John. And that’s all that’s needed for her to suffer from illusions of meaning, assertion, and thought.

<sup>14</sup> See his (2013).

We think that Cases (1)-(3) are compelling examples of deceptive nonsense and illusions of thought. Some philosophers have argued, however, that illusions of thought are *impossible*. Even though we are ultimately unpersuaded by these arguments, we think they have been underestimated. We focus here on a recent discussion by Herman Cappelen: in this section, we argue that Cappelen's responses to these arguments are unsatisfying, and in §5 we'll give what we take to be a more convincing response to them.

**4.1 The Original Argument.** Cappelen presents the first argument as follows:<sup>15</sup>

### The Original Argument

- O1. Illusion requires falsity. The thinker must have a false belief (or other attitude) about her propositional attitudes.
- O2. A natural development of (1) is that for a subject to have the illusion of thought, she must have a false belief of the form *I was thinking that p*.
- O3. Suppose *p* is nonsense.
- O4. Then: *I was thinking that p* can't be nonsense, since it has to be false.
- O5. Step 4 requires an account of (or semantics for) the second-order thought that allows the complement *p* to be nonsense and the entire self-attribution to be not-nonsense and false.
- O6. The correct account or semantics for second-order thoughts requires that the complement in *A thinks that p* to be propositional.
- O7. So: illusion of thought is impossible. (Cappelen 2013: 29)

**4.2 Cappelen's Objections.** Cappelen's discussion of The Original Argument is brief and clear, so we reproduce it here with minimal editorializing. He says that (O1)

is dubious or at least in need of further argument: we could think of an illusion of thought along the lines of an illusion of a dagger, where that is not to be construed as having a false belief about the presence of a dagger, but simply as what it looks like grammatically: the illusion of a dagger. Similarly, we can have the illusion of a thought and not construe that as the having of a false belief about a thought-like event. (2013: 30)

He says that (O2)

is dubious because even if you think illusions of thought require false beliefs about thoughts, the false thoughts need not be of the form, *I was thinking that p*. It could be a demonstrative thought of the form, *That was a thought* (accompanied by a demonstration of the cognitive event that was not a thought). A demonstrative thought of that form would be false if the demonstrated event wasn't a thought (i.e. we have no reason to think the demonstrative thought is nonsense just because it demonstrates nonsense). (2013: 29)

Finally, Cappelen says that (O6) is dubious given that

there's no consensus on what the correct semantics for belief reports is (and no consensus on the correct account of second-order thought), and so any claim about what the correct

---

<sup>15</sup> According to Cappelen, this argument was inspired by an argument Paul Boghossian presented (but didn't endorse) in conversation.

semantics *allows* will be controversial and require very substantive theoretical commitments...Putting that lack of consensus aside, I know of no view that rules out an account of second-order thought according to which such a thought *presupposes* that the complement is propositional, and if that presupposition fails, the thought is false. (2013: 29-30)

**4.3 The Revised Argument.** We think that Cappelen's objections are fairly decisive against The Original Argument. But we also think that The Original Argument is unnecessarily weak, and that there's a version of the argument that is simpler, more intuitively compelling, and less susceptible to Cappelen's objections. It runs as follows (where '*p*' stands for a sentence or sentence-like linguistic string):

### The Revised Argument

R1. If it's possible for a subject *S* to have the illusion of thinking that *p* in context *c*, then it's jointly possible that, in *c*, *p* lacks content and *S* falsely (or truly) says or thinks something of the form *I was thinking that p*. ( $\Diamond I \supset \Diamond(L \bullet (S \vee B))$ )

R2. It's not jointly possible, in a single context *c*, for *p* to lack content and for *S* to falsely (or truly) say or think something of the form *I was thinking that p*. ( $\sim\Diamond(L \bullet (S \vee B))$ )

R3. So, it's not possible for a subject *S* to have the illusion of thinking that *p*. ( $\sim\Diamond I$ )

The Revised Argument seems to straightforwardly avoid Cappelen's first two criticisms of The Original Argument. In short, it does this by changing the modality of the claims from necessity to possibility. Contra (O1) (and in line with Cappelen's critique of it), The Revised Argument grants that not all illusions of thought are *necessarily* accompanied by false second-order thoughts about them: one *might* have an illusion of thought without having a false (or true!) second-order belief about it. Instead, (R1) merely insists that, given that there are illusions of thought, it's *possible* to have false (or true) second-order beliefs about them, and in particular second-order beliefs of the form *I was thinking that p* (and likewise for one to falsely or truly say something of the form *I was thinking that p*.)

Similarly, contra (O2) (and in line with Cappelen's critique of it), The Revised Argument grants that false (or true) second-order beliefs (and their verbal expressions) about the cognitive episodes that are illusions of thought *needn't* be of the form *I was thinking that p*. Rather, (R1) merely insists that, given that there are illusions of thought, it's *possible* to have false (or true) second-order beliefs about those episodes that *are* of the form *I was thinking that p* (and to report those beliefs with sentences of the form *I was thinking that p*).

The Revised Argument does *not* avoid Cappelen's critique of (O6), since that critique applies equally well to (R2). But applying equally well does not entail applying well, and it's not clear that Cappelen's critique applies well to either (O6) or (R2). While it's true that there's no *consensus* about the semantics of belief reports (including second-order thoughts), and that not *every* going theory requires that-clauses embedded in belief reports to be meaningful for the belief-reports themselves to be

meaningful, it's also true that many of the most popular theories do. And that seems intuitively correct. But then, as David Bell writes,

The difficulty is, crudely speaking, that either the non-existence of the embedded, merely apparent thought will contaminate the second-order thought of which it is a part, or, conversely, the intelligibility of the second order thought will bestow respectability on its first order component. (1988: 51)

These considerations seem sufficient for making The Revised Argument interesting and important, if not successful. We don't claim that The Revised Argument is knockdown; after all, we reject its conclusion. But if resisting The Revised Argument forces us to reject many, or even *any*, going theories about second-order thought and the semantics of belief reports, that's enough to give it some bite.

**4.4 The Action-Explanation Argument.** Cappelen also considers a line of argument against illusions of thought presented in Segal (2000), Wikforss (2007), and O'Brien (2009). In a nutshell, the argument is that illusions of thought would undermine our ability to explain certain aspects of agents' behavior. As Segal puts it:

The main argument for attributing empty concepts [as the contents of expressions that defenders of illusions of thought would say have no content] in all these cases...is simply that by so doing, and only by so doing, can we make psychological sense of a very wide variety of human activity and cognition. (2000: 37)

Along similar lines, Wikforss says:

From the point of view of the individual, after all, it is as if there was a thought available, one that she reasons with and acts on...How can this be explained if one endorses [the claim that these "thoughts" are mere illusions]? (2007: 173)

And O'Brien writes,

Even when A fails to suppose that *P*, due to content failure, it seems to her that she supposes that *P*, and she can act and infer as she would, were she...supposing that *P*. On the gap view we have no explanation of why it seems to A that she is supposing that *P*, or of her actions consequent on its seeming to her she is supposing 'that glass is heavy', or of her inferring that 'if my supposition is true, there is at least one heavy glass'. We seem to need *something* to do the normative and epistemic work—some associated act, or some remnant or degraded version of the act one gets in the good case. (2009: 219)

In response to what we'll call **The Action-Explanation Argument**, Cappelen says that

...when there's illusion of thought, there is illusion of reasoning, and so illusion of practical reasoning. It should come as no surprise that we are fallible with respect to the nature of the cognitive mechanisms that trigger action. (2013: 31)

He goes on to point out that this doesn't rule out explaining our behavior by appeal to our beliefs, desires, and reasoning:

Such an explanation could, at least in principle, appeal to the illusion of thought, i.e. that there was an illusion can be part of the explanation. It can also...appeal to second-order thoughts about the nonsensical cognitive event. Here is [a story about why someone would (sincerely) write a nonsensical sentence]: *Martin thought that that (demonstrating some cognitive event) was a*

*thought, he wanted to communicate that thought to others, he thought that the sentence S expressed that thought and that by writing down S his desire could be fulfilled.* So, he had a bunch of false second-order beliefs, and those, combined with his desires, explain his actions. (2013: 31)

**4.5 Contra Cappelen on The Action-Explanation Argument.** Our concern with this response to The Action-Explanation Argument isn't that the argument can be reformulated so as to avoid it. Rather, our concern is that it's implausible to explain actions ostensibly arising from illusions of thought by appeal to second-order thoughts about those illusions. First, as noted above, we just don't think that we engage in that much second-order thought: we don't think the second-order thoughts required by this explanation are always *there*. Second, it's hard to see how the sort of demonstrative second-order thoughts countenanced by Cappelen (e.g., *that was a thought*) could be sufficient to explain one's behavior: to explain why Martin said 'das Nichts nichtet' (rather than, say, 'Witches cast spells'), it's not enough to point out that he had the second order thought *that was a thought*. For a second-order thought to do the job, it seems like it needs to be a thought of the form *I was thinking that das Nichts nichtet*.<sup>16</sup> But it's unclear how such second-order thoughts could be meaningful if 'das Nichts nichtet' isn't. (Recall our discussion of (R2) above.) Finally, Cappelen's response involves a kind of disjunctivism, treating the springs of action as categorically different when they are illusions of thought: in "good" cases, first-order thoughts bring about actions (in conjunction with our desires), but in illusion-of-thought cases, second-order thoughts do all the work. While we are happy to grant that there are illusions of practical reasoning and that we don't always know the reasons why we do the things we do, it seems implausible that the role of illusions of thought in our mental economy is so different than that of proper thoughts: if so, why are such illusions so hard to detect? At the very least, it seems, this hypothesis would predict some sort of experimentally measurable difference in cognitive functioning when our thoughts are nonsensical, since we would have to abandon our normal *modus operandi* and fall back on our second-order thoughts (assuming they exist). While we are hesitant to endorse empirical claims from the armchair, we don't find the existence of such a difference plausible.<sup>17</sup>

## 5. RESOLVING THE ANTINOMY

We think that the considerations presented in §3 (and the cases in §2) fairly conclusively establish the existence of deceptive nonsense and illusions of thought. But we also think that the arguments in §4 for the *impossibility* of illusions of thought are intuitively compelling, and that Cappelen's objections to them are unsatisfying. In this section, we explain how to resolve this apparent antinomy. We think the most satisfying resolution appeals to the Language of Thought Hypothesis (LOT), according to which

---

<sup>16</sup> Cappelen says that Martin also believes that 'das Nichts nichtet' expresses the thought he just had. But *why* does Martin think this? How do you get from the belief that *that was a thought* to the belief that 'das Nichts nichtet' expresses that thought? Presumably that's not just a brute fact, but the demonstrative second-order thought could be referring to anything, and so it's hard to see what leads Martin to choose the words he does.

<sup>17</sup> See O'Brien (2009: §3.3) for further discussion of problems with disjunctivist responses to The Action-Explanation Argument.

thought occurs in a (non-conventional) mental language (“Mentalese”) with its own syntax and semantics.<sup>18</sup> But we think a similar, if slightly less satisfying, resolution strategy is available to those who reject LOT.

**5.1 Thinking Nothing.** To set the stage for what follows, consider a final mystery confronting defenders of illusions of thought: explaining how it could be that “nothing at all” is going on when we talk, think, and write nonsense. Carnap, for example, says,

...how could it be explained that so many men in all ages and nations, among them eminent minds, spent so much energy, nay veritable fervor, on metaphysics if the latter consisted of nothing but mere words, nonsensically juxtaposed? And how could one account for the fact that metaphysical books have exerted such a strong influence on readers up to the present day, if they contained not even errors, but nothing at all? (1959: 78)

And Cappelen writes,

[Wittgenstein and Carnap] thought that many of those we consider great thinkers were not thinking at all. According to Carnap, what some considered the high points of human intellectual achievement are no more than a bunch of people making noises and marks on paper. Those who read, commented on, and developed their work suffered from the same illusion. They had what appear to be discussions; they wrote books and papers apparently responding to each other. But it was all the most fundamental kind of failure: it was neither true nor false, no thoughts were expressed, and there was no agreement or disagreement. It was all just a complete waste of time, energy, ink, and paper. (2013: 23)

**5.2 Thinking Empty Thoughts.** We think that these passages are at least potentially misleading, and the mystery they raise about illusions of thought is itself an illusion. There is no need to say that those who “think nonsense” are not really thinking: all we’re committed to is saying that their thinking lacks content. That is, we can simply deny what O’Brien (2009: 215) calls The Dependence of Thought on Content Thesis, according to which there is no thinking without content.<sup>19</sup>

Words like ‘belief’ and ‘thought’ can be used to refer both to the *act* of believing or thinking something (a concrete mental state or event), and to the *contents* of such states (abstract propositions). When we talk about *sharing beliefs*, we are (usually) talking about numerically identical belief *contents* (propositions) shared by people in numerically distinct belief *states*. So ‘beliefs’ sometimes refers to propositions believed—the *contents* of our belief states—but it can also refer to belief states themselves, as when we say things like “My belief that John is abominable was influenced by yours”. To keep things clear in what follows, we’ll use ‘beliefs<sub>s</sub>’ to refer to states of believing and ‘beliefs<sub>c</sub>’ to the contents (if any) of such states.

---

<sup>18</sup> Fodor (1975) is the *locus classicus* of contemporary discussion of LOT, although our proximate inspiration is Braun (1993). Similar views were prominent in medieval philosophy; see, e.g., Karger (1996) and Bulthuis (2020).

<sup>19</sup> O’Brien claims that this principle is widely endorsed by defenders of illusions of thought.

With this distinction in hand, we can say that while illusions of thought do not, by definition, involve thought *contents*, they may still involve beliefs<sub>s</sub> (and hopes<sub>s</sub> and desires<sub>s</sub>). It's just that those beliefs<sub>s</sub>, hopes<sub>s</sub>, and desires<sub>s</sub>—those thoughts<sub>s</sub>—are *empty*.<sup>20</sup> We can thus give a unified account of thinking, speaking, and writing nonsense: thinking nonsense involves *actually* thinking empty thoughts; speaking nonsense involves *actually* speaking empty words, and writing nonsense involves *actually* writing sentences that don't (actually) express contents. Such writers aren't writing *nothing*—they might be very productive—it's just that the writing they produce lacks content. On this picture, thinking, speaking, and writing nonsense is clearly possible. And contrary to what the above quotes suggest, it isn't even always a waste of time: thinking, writing, and speaking nonsense on the way to formulating some important truth (or falsehood) wouldn't be. But even when it *is* a waste of time, it isn't doing *nothing*.

**5.3 The Language of Thought.** According to one way of thinking about beliefs<sub>s</sub> and their relation to beliefs<sub>c</sub>, to have a belief<sub>s</sub> is to have a Mentalese sentence in one's "belief box",<sup>21</sup> and to have a belief<sub>c</sub> is to have a Mentalese sentence in one's belief box that has that belief<sub>c</sub> as its content, a situation we will sometimes describe as the belief<sub>s</sub> having the belief<sub>c</sub> as its content. It is straightforward to consistently describe illusions of thought on this view: in Cases (1)-(3), 'John is abominable' is semantic nonsense, and Maeve has a Mentalese translation of 'John is abominable' ('JOHN IS ABOMINABLE') in her belief box, a translation that is also nonsense. So 'John is abominable' is (deceptive) sentential nonsense, Maeve's utterance of 'John is abominable' is (deceptive) assertoric nonsense, Maeve's belief<sub>s</sub> is (deceptive) cognitive nonsense. Maeve is thus suffering from an illusion of thought.

Nonetheless, Maeve's utterance of 'John is abominable' was (partially) caused by one of her (first-order) beliefs<sub>s</sub>. Indeed, there's a sense in which it was (partially) caused by her belief<sub>s</sub> that John is abominable. After all, it was caused by 'JOHN IS ABOMINABLE' being in her belief box. That's true even though Maeve doesn't have the belief<sub>c</sub> that John is abominable, since there *is* no such belief<sub>c</sub>, and *a fortiori* no such belief<sub>c</sub> expressed by 'JOHN IS ABOMINABLE'. We can say, then, that illusions of thought involve real beliefs<sub>s</sub>: real sentences in one's belief box. It's just that those beliefs/sentences don't have contents. That's the illusion: you are deceived about your belief<sub>s</sub> having content. But you are not deceived about the existence of the belief<sub>s</sub> itself. From this perspective, illusions of thought mislead us about the existence of beliefs<sub>c</sub> but not the existence of beliefs<sub>s</sub>.

**5.4 Unmediated Belief.** According to an alternative conception of beliefs<sub>s</sub> and their relation to beliefs<sub>c</sub>, a belief<sub>s</sub> is simply a relation between a subject and proposition (the relevant belief<sub>c</sub>), unmediated by any sort of mental sentence. On this view, if there's no proposition (no belief<sub>c</sub>), there's

---

<sup>20</sup> O'Brien herself seems to endorse something like distinction. She writes, "Given the possibility of a similar such ambiguity in the case of thought, [one] can very reasonably suggest that the thinking involved in rehearsing a contentless syntactic string is not the kind of thinking at issue in the Dependence of Thought on Content Thesis." (2009: 228)

<sup>21</sup> See Schiffer (1981). This is obviously a metaphor: having a sentence in one's "belief box" is supposed to be something like the mental analogue of assertion: something like assent. Merely having a Mentalese sentence "in one's head" is insufficient for belief, since there are other propositional attitudes one might have towards its content: one might hope that *p*, fear that *p*, wonder whether *p*, and so on.

no belief<sub>s</sub>. Hence, when one suffers from an illusion of thought, there's no belief there *at all*—not just no belief<sub>s</sub>, but no belief<sub>s</sub> either: nothing analogous to having a Mentalese sentence in one's belief box. Of course, one is presumably in a mental state similar to the state one is in when one believes something: that's what generates the illusion. But that mental state isn't a belief, and so, strictly speaking, illusions of thought don't involve beliefs in either the act or content sense, but merely illusions of belief. On this view, illusions of thought mislead us about the existence of both beliefs<sub>s</sub> and beliefs<sub>e</sub>. Perhaps those drawn to O'Brien's Dependency of Thought on Content Thesis are presupposing a view like this. Note, however, that this unmediated picture of belief doesn't entail that there's *nothing* going on when one is having an illusion of thought: there's still a belief-*like* state that one is in, even if that state isn't actually a belief, due to its lack of content.

**5.5 Belief Reports.** To see how these considerations allow us to respond to the arguments from §4, note that there are three ways of thinking about **nonsense reports** (statements of the form *S believes that p*<sup>22</sup> in cases where *p* lacks content, such as 'Maeve believes that John is abominable'): they might be considered true, false, or nonsense (neither). Here are sets of truth-conditions illustrating these possibilities for both LOT and Unmediated views of belief:

	LOT	Unmediated
False	(LF) 'Maeve believes that John is abominable' is true iff the proposition expressed by 'John is abominable' is expressed by some sentence in Maeve's belief box.	(UF) 'Maeve believes that John is abominable' is true iff Maeve stands in the relation of believing to the proposition expressed by 'John is abominable'.
Nonsense	(LN) 'Maeve believes that John is abominable' is true iff Maeve has a Mentalese sentence <i>m</i> in her belief box such that [[ <i>m</i> ]] = [['John is abominable']].	(UN) 'Maeve believes that John is abominable' is true iff Maeve stands in the relation of believing to [['John is abominable']].
True	(LT) 'Maeve believes that John is abominable' is true iff Maeve has a Mentalese sentence <i>m</i> in her belief box that is an adequate translation of 'John is abominable'.	(UT) 'Maeve believes that John is abominable' is true iff Maeve is disposed to accept or assent to 'John is abominable'.

Given relevant plausible background assumptions (including a Russellian treatment of definite descriptions), LF and UF would make 'Maeve believes that John is abominable' false, since there is no proposition satisfying the description on their right-hand sides. LN and UN, on the other hand, would

<sup>22</sup> Including second-order thoughts of the form *I believe that p*.

make the nonsense report itself nonsense, since if ‘John is abominable’ has no content, then [[‘John is abominable’]] doesn’t exist, which makes the right-hand-sides of the biconditionals in LN and UN nonsense (neither true nor false), which makes the left-hand-sides nonsense (neither true nor false).<sup>23</sup> For obvious reasons, LT and UT would make the nonsense report true.

**5.6 The Revised Argument Revisited.** What, then, about The Revised Argument? Consider this instance of it:

R1. If it’s possible for Maeve to have the illusion of thinking that John is abominable (in context *c*), then it’s jointly possible that, in *c*, ‘John is abominable’ lacks content and Maeve falsely (or truly) thinks or says something of the form *I was thinking that John is abominable*.

R2. It’s not jointly possible, in any context (including *c*), for ‘John is abominable’ to lack content and for Maeve to falsely (or truly) think or say something of the form *I was thinking that John is abominable*.

R3. So, it’s not possible for Maeve to have the illusion of thinking that John is abominable.

If nonsense reports are *false*, then (R1) is true but (R2) is false. If, however, nonsense reports are *nonsense*, then (R1) is false but (R2) is true. Finally, if nonsense reports are *true*, then (R1) is true but (R2) is false. There is, then, no way of assigning truth-conditions to nonsense reports makes both premises of The Revised Argument true.<sup>24</sup> This is happy news, since defending some particular set of truth conditions for nonsense reports would require a paper unto itself, and we’re past our word limit. But if there is *no* way of assigning truth-conditions to nonsense reports that makes both premises of The Revised Argument true, we can conclude without further ado that it is unsound. It is worth noting, however, that each of The Revised Argument’s premises is vindicated by *some* (vaguely plausible) way of assigning truth conditions to nonsense reports. This might explain the intuitive appeal of the argument.

**5.7 The Action-Explanation Argument Revisited.** Now let’s reconsider the problem of explaining actions based on illusions of thought. From the perspective of LOT, the solution is straightforward: Maeve utters ‘John is abominable’ because of the presence of ‘JOHN IS ABOMINABLE’ in her belief box, its lack of content notwithstanding. There is no need to appeal to second-order thoughts at all, and hence no need to appeal to them too extensively. And the presence of ‘JOHN IS ABOMINABLE’ in Maeve’s belief box seems like a significantly *better* explanation of her utterance of ‘John is abominable’ than the second order thought that *that was a thought*, even if she has that second-order thought. In general, the presence of nonsense beliefs<sub>s</sub> (nonsense Mentalese sentences in one’s belief box) allows us to explain how nonsense functions in our mental economy without constant appeal to second-order thought: nonsense thoughts<sub>s</sub> (beliefs<sub>s</sub>, desires<sub>s</sub>, etc.) exist, and so can play their normal role as springs of action, even though they have no content.

---

<sup>23</sup> At least if ‘[[‘John is abominable’]]’ has a Millian semantics. The example would need to be modified in a Fregean context.

<sup>24</sup> Hence, resisting the argument doesn’t require us to reject *any* view about the truth-conditions of belief reports.

Things are more complicated when it comes to the unmediated picture of belief, according to which illusions of thought don't involve *any* (relevant) beliefs<sub>s</sub> or beliefs<sub>c</sub>. Hence, we can't say that Maeve's beliefs<sub>s</sub> are playing their normal role in bringing about her behavior. The obvious fallback position is to appeal to the belief<sub>s</sub>-like cognitive states—*failed* beliefs<sub>s</sub>—that give rise to Maeve's illusions of thought in our explanation of Maeve's speech and deeds. But since these belief<sub>s</sub>-like cognitive states will not be beliefs<sub>s</sub>, this story will be less elegant and straightforward than the one that adherents of LOT are able to give. When it comes to the Action-Explanation Argument, then, LOT theories seem to have an advantage over unmediated conceptions of belief.

**5.8 O'Brien's Objection.** Are there objections to LOT theories that blunt this advantage? Lucy O'Brien (2009) gives an argument against a LOT-like account,<sup>25</sup> contending that it doesn't provide a satisfactory account of the cognitive significance of illusions of thought. She begins by considering a case involving the perception of two similar glasses, where one engages in distinct episodes of successful demonstrative thought about each, which the LOT theorist would gloss as 'THAT GLASS IS HEAVY' appearing twice in one's belief box. (Call these two token Mentalese sentences T1 and T2.) Despite their lexical similarity, the reason T1 and T2 count as two beliefs, rather than one, is that their *contents* are different: those Mentalese sentences have different truth-conditions, since the tokens of 'THAT' that appear in them denote different glasses.

O'Brien goes on to consider a phenomenologically indistinguishable case where one merely *hallucinates* the two glasses, and hence where there is no demonstrated content to distinguish T1 and T2. Nonetheless, "it seems to [one] that she is thinking two thoughts and it seems to her that they are distinct thoughts." (2009: 227) Here, T1 and T2 have no content, and so cannot be distinguished by their contents. Hence, it seems that LOT cannot explain the fact that the subject is suffering from two illusions of thought, rather than one.

It's worth noting that a similar problem might arise for names: if Maeve knows *two* horrible people named 'John', she might have two tokens of the LOT sentence 'JOHN IS ABOMINABLE' in her belief box, which "count as" two beliefs since those instances of 'JOHN' refer to different people (and so have different contents). But Maeve could also suffer from two illusions of thought she would express with the sentence 'John is abominable'. In that case, we could not distinguish the LOT sentences in her belief box by their contents, since *ex hypothesi*, they don't have any. Thus, LOT does not seem to be able to explain the fact that Maeve is suffering from two illusions of thought, rather than one.

**5.9 The Indexing Response.** The response to both the demonstrative and nominal versions of this objection is, we think, the same: indexing. If names are *singular* terms, they cannot refer to more than one thing. Hence, it is common to think that there is not one English name, 'John', but indefinitely

---

<sup>25</sup> Essentially a "silently rehearsing sentences to oneself" view. In what follows, we have adapted her argument to apply directly to LOT.

many homonymous names spelled J-O-H-N.<sup>26</sup> If that's right, we should presumably say the same thing about LOT names: we don't have two instances of 'JOHN', but one instance of 'JOHN<sub>1</sub>' and another of 'JOHN<sub>2</sub>'. There is obviously no problem distinguishing 'JOHN<sub>1</sub> IS ABOMINABLE' from 'JOHN<sub>2</sub> IS ABOMINABLE'.

A similar "indexing approach" to natural-language demonstratives is endorsed by Kaplan, and by Fodor when it comes to Mentalese demonstratives.<sup>27</sup> If this is correct, the response to O'Brien's demonstrative case is the same as it was in the case of names: 'THAT<sub>1</sub> GLASS IS FULL' and 'THAT<sub>2</sub> GLASS IS FULL' are distinct LOT sentences (composed of distinct LOT words), and so there's no need to appeal to their (missing) contents to distinguish them.<sup>28</sup> And if that's right, LOT theories maintain their advantage over unmediated conceptions of belief with regard to the Action-Explanation Argument.

## 6. CONCLUSION

We have argued that while deceptive nonsense and illusions of thought exist, they are not as bad as they might seem. Some authors suggest that illusions of thought involve ignorance of what is going on in our heads, of what mental states we are in and whether we are even thinking at all. While we're as pessimistic as anyone about the extent of human ignorance, we have argued that *this* sort of ignorance is not, in fact, entailed by illusions of thought: such objections to illusions of thought are much ado about nothing. We *do* know we're thinking. We might even know *what* we're thinking. We just don't always know what, if anything, we are thinking about.<sup>29</sup>

## BIBLIOGRAPHY

- Ayer, A. J. 1952. *Language, Truth, and Logic*, 2<sup>nd</sup> edition. New York City: Dover.
- Baier, Annette. 1972. 'Nonsense'. In *The Encyclopedia of Philosophy* v.5, Macmillian, edited by P. Edwards.
- Bell, David. 1988. 'Phenomenology, Solipsism, and Egocentric Thought'. *Proceedings of the Aristotelian Society, Supplementary*, vol. 62: 45-60.
- Bentham, Jeremy. 1843. 'Anarchical Fallacies'. *The Works of Jeremy Bentham* v.2, edited by J. Bowring. Edinburgh: William Tait.
- Braun, David. 1993. 'Empty Names'. *Noûs* vol. 27 (no. 4.) (Dec. 1993): 449-469.

---

<sup>26</sup> For example, Kripke says that "uses of phonetically the same sounds to name distinct objects count as distinct names", at least "for theoretical purposes". (1980: 8) This view is endorsed in Kaplan (1990), and elsewhere.

<sup>27</sup> See, e.g., Kaplan (1989), and Stojnić and Lepore (2020) for discussion.

<sup>28</sup> Stojnić and Lepore 2020 raise significant objections to the Kaplan-Fodor strategy, but they ultimately offer a response to those objections on Fodor's behalf that is friendly to our appeal to LOT in explaining the cognitive episodes underlying illusions of thought.

<sup>29</sup> To avoid any confusion, we intend the previous three sentences to communicate the following claims: we (reliably) know *whether* there are sentences in our belief, desire, etc. boxes; we might even (reliably) know *what* sentences are in our belief, desire, etc. boxes; but we less reliably know what, if any, *contents* those sentences have. Thanks to Audre Brokes, Nathaniel Bulthuis, Herman Cappelen, Joseph Corabi, Todd Moody, Andrew Payne, and Brice Wachterhauser for helpful comments and discussion.

- Braun, David. 2015. 'Wondering About Witches'. *Fictional Objects*, edited by A. Everett and S. Brock, Oxford: Oxford University Press.
- Bulthuis, Nathan. 2020. 'Walter Burley on Co-Signification in Opaque Contexts'. In *Oxford Studies in Medieval Philosophy*, edited by Robert Pasnau, Oxford: Oxford University Press.
- Cappelen, Herman. 2013. 'Nonsense and Illusions of Thought'. *Philosophical Perspectives: Philosophy of Language* vol. 27: 22-50.
- Cappelen Herman, and Lepore, Ernest. 2005. *Insensitive Semantics*, Oxford: Blackwell.
- Carnap, Rudolph. 1959. 'The Elimination of Metaphysics through Logical Analysis of Language'. In *Logical Positivism*, edited by A. J. Ayer, translated by Arthur Pap. New York: The Free Press.
- Diamond, Cora. 1981. 'What Nonsense Might Be'. *Philosophy* vol. 56 (no. 215): 5-22.
- Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.
- Fodor, Jerry. 1975. *The Language of Thought*. Cambridge: Harvard University Press.
- Frege, Gottlob. 1956. 'The Thought: A Logical Inquiry'. *Mind* vol. 65 (no. 259) (July 1956): 289-311.
- Heck, Richard. 2002. 'Do Demonstratives Have Senses?'. *Philosopher's Imprint* vol. 2 (no. 2) (June 2002): 1-33.
- Heidegger, Martin. 1929. 'What Is Metaphysics'. Freiburg Inaugural Lecture.
- Kaplan, David. 1989. 'Demonstratives'. In *Themes from Kaplan*, edited by in Joseph Almog, John Perry, and Howard Wettstein. New York: Oxford University Press.
- Kaplan, David. 1990. 'Words.' *Proceedings of the Aristotelian Society, Supplementary Volumes* vol. 64: 93-119.
- Karger, Elizabeth. 1996. 'Mental Sentences According to Burley and to the Early Ockham.' *Vivarium* vol. 34 (no. 2): 192-230.
- Keller, John A. 2017. 'Introduction'. In *Being, Freedom, and Method: Themes from the Philosophy of Peter van Inwagen*, edited by John A. Keller. Oxford: Oxford University Press.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.
- Magidor, Ofra. 2013. *Category Mistakes*. Oxford: Oxford University Press.
- McDowell, John. 1986. 'Singular Thought and the Extent of Inner Space'. In *Subject, Thought, and Context*, edited by Phillip Pettit and John McDowell. New York: Oxford University Press.
- O'Brien, Lucy. 2009. 'Mental Actions and the No-Content Problem'. In *Mental Actions*, edited by in Lucy O'Brien and Matthew Soteriou. Oxford: Oxford University Press.
- Pryor, James. 2006. 'Hyper-reliability and Apriority'. *Proceeding of the Aristotelian Society* vol. 106 (1): 329-346.
- Quine, W.V.O. 1960. *Word and Object*. Cambridge: Harvard University Press.
- Recanati, Francois. 1997. 'Can We Believe What We Do Not Understand?'. *Mind & Language* vol. 12 (1): 84-100.
- Russell, Bertrnd. 1956. 'The Philosophy of Logical Atomism'. In *Logic and Knowledge*, edited by Robert Marsh. London: George Allen & Unwin.
- Schiffer, Stephen. 1981. 'Truth and the Theory of Content'. In *Meaning and Understanding*, edited by Herman Parret and Jacques Bouveresse. New York: Walter de Gruyter.
- Segal, Gabriel. 2000. *A Slim Book About Narrow Content*. Cambridge: MIT Press.
- Soames, Scott. 2005. 'Naming and Asserting'. In *Semantics vs. Pragmatics*, edited by Zoltan Szabó. Oxford: Oxford University Press.
- Stojnić, Una and Lepore, Ernie. 2020. 'Fodor and demonstratives in LOT'. *Theoria* vol. 35 (1): 75-92.
- Van Inwagen, Peter. 1980. 'Philosophers and the Words "Human Body"'. In *Time and Cause: Essays Presented to Richard Taylor*, edited by Peter van Inwagen. Dordrecht: D. Reidel Publishing Company.
- Van Inwagen, Peter. 2017. 'Concluding Meditation'. In *Being, Freedom, and Method: Themes from the Philosophy of Peter van Inwagen*, edited by John A. Keller. Oxford: Oxford University Press.
- Wikforss, Åsa. 2007. 'Semantic Externalism and Psychological Externalism'. *Philosophy Compass* vol. 3 (1): 158-181.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. New York: Oxford.

Wittgenstein, Ludwig. 1961. *Tractatus Logico-Philosophicus*, translated by D. F. Pears and B. F. McGuinness. London: Routledge and Kegan Paul.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*, translated by G. E. M. Anscombe. Oxford: Basil Blackwell.