ABSTRACT

Title of dissertation: SEMANTICS AND PRAGMATICS IN A MODULAR MIND

Michael McCourt
Doctor of Philosophy, 2021

Dissertation directed by: Professor Alexander Williams
Department of Linguistics

This dissertation asks how we should understand the distinction between semantic and pragmatic aspects of linguistic understanding within the framework of mentalism, on which the study of language is a branch of psychology. In particular, I assess a proposal on which the distinction between semantics and pragmatics is ultimately grounded in the modularity or encapsulation of semantic processes. While pragmatic processes involved in understanding the communicative intentions of a *speaker* are non-modular and highly inferential, semantic processes involved in understanding the meaning of an *expression* are modular and encapsulated from top-down influences of general cognition. The encapsulation hypothesis for semantics is attractive, since it would allow the semantics-pragmatics distinction to cut a natural joint in the communicating mind.

However, as I argue, the case in favor of the modularity hypothesis for semantics is not particularly strong. Many of the arguments offered in its support are unsuccessful. I therefore carefully assess the relevant experimental record, in

rapport with parallel debates about modular processing in other domains, such as vision. I point to several observations that raise a challenge for the encapsulation hypothesis for semantics; and I recommend consideration of alternative notions of modularity. However, I also demonstrate some principled strategies that proponents of the encapsulation hypothesis might deploy in order to meet the empirical challenge that I raise. I conclude that the available data neither falsify nor support the modularity hypothesis for semantics, and accordingly I develop several strategies that might be pursued in future work.

It has also been argued that the encapsulation of semantic processing would entail (or otherwise strongly recommend) a particular approach to word meaning. However, in rapport with the literature on polysemy—a phenomenon whereby a single word can be used to express several related concepts, but not due to generality—I show that such arguments are largely unsuccessful. Again, I develop strategies that might be used, going forward, to adjudicate among the options regarding word meaning within a mentalistic linguistics.

SEMANTICS AND PRAGMATICS IN A MODULAR MIND

by

Michael Sullivan McCourt

Dissertation submitted to the Faculty of the Graduate School of
the University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:

Professor Alexander Williams (Chair)
Professor Peter Carruthers
Professor Yi Ting Huang (Dean's Representative)
Professor Ellen Lau
Professor Paul Pietroski
Professor Georges Rey

Dedicated to Emma
and *uncer giedd geador* (our song together)

Acknowledgements

I owe thanks here, first, to everyone with and from whom I have learned about philosophy and linguistics over the years. As usual, while the faults in what follows are mine, any virtues in this are largely due to the village.

This dissertation is a product of the conditions that have fostered and sustained a vibrant community of people studying language in College Park, Maryland. Thanks are due here in particular to participants in reading groups and lab meetings on campus who have offered feedback at many stages of this project: the Cognitive and Neuroscience Lab (CNL), Language Science Lunch Talks (LSLT), the Philosophy and Linguistics reading group (PHLING), Philosophy colloquia, the Semantics and Language Acquisition Lab (SLAB), Works in Progress (WiP), and others. Thanks to all who helped me work through these ideas at various stages, offering feedback and encouragement along the way.

An NSF fellowship allowed me, during pre-candidacy research, to take a few years off from teaching, pursue extra coursework in linguistics, and collaboratively run several psycholinguistic experiments. The IGERT years resulted in some conferences and publications, but also a deeper appreciation for what's involved in bringing philosophy and linguistics into conversation with each other. That appreciation laid important groundwork for the research project described in what follows here. Many thanks are also due to the co-authors and co-experimenters I started working with during the IGERT years: Ellen Lau, Alexander Williams, and Jeff Green. I learned a lot from our projects together.

I can't mention by name all of the teachers and fellow graduate students who are owed some thanks here, not without missing a few. But that's what quantifiers are for. To all my teachers and colleagues over the years—along the banks of the Connecticut, the Kishwaukee, and the Anacostia—thank you. Getting here without you all would have been, at best, much harder. To the gang at Rushmore at NIU, Queensbury House at UMD, and various affiliates and satellites: I've been lucky to study alongside such great people.

Extra special thanks are due to my colleague, Quinn Harr—longtime officemate, fellow advisee under Alexander, and a willing and resilient sparring partner, both philosophically and on the tennis court. Thanks for going ahead of me, for guiding me along the route, and for being a dear friend throughout.

Among my more remote pedagogical ancestors, I'd like to specifically thank a few representatives. Thanks to Anthony Rizzitano, who was first my high school English teacher and then later a close friend, and who taught me how to properly read a piece of fiction and write a decent essay (and also how to paint the exteriors of coastal New England houses by hand, with no power tools). I'd also like to thank Adam Zucker (Massachusetts, Amherst) and James Harold (Mount Holyoke College), whose mentorship and encouragement at the end of one stage me to the next.

At Northern Illinois University, special thanks are owed to David Buller, Carl Gillett, and Geoff Pynn. Aside from the UMD faculty on my thesis committee, I'd also like to thank Rachel Singpurwalla, who has been a mentor when it comes to my own philosophical pedagogy. Relatedly, I should also thank

the hundreds of UMD undergraduate students with whom I have talked about words and concepts and other things over the years.

Thanks to the members of my committee—Alexander Williams, Peter Carruthers, Yi Ting Huang, Ellen Lau, Paul Pietroski, and Georges Rey. Thank you all for reading this document and offering lots of helpful feedback during the oral defense.

I'd like to thank Paul Pietroski for writing the articles about words and concepts that brought me here to UMD in the first place, for the many helpful conversations we had as I developed this project, and for feedback on the final product. Thanks to Ellen Lau for mentorship and collaboration in and out of the lab, for many enlightening conversations about meaning and use, and more recently for helpful feedback on drafts of this dissertation. Thanks to Peter Carruthers for his professional and pedagogical mentorship, for conversations and seminars that shaped my understanding of cognitive science, and for advice that was formative in the development of my ideas about meaning and modularity. Lastly, special thanks are owed to Georges Rey, for conversations and guidance over the years, for teaching me a lot about Chomsky and Fodor, for comradery at R.J. Bentley's, and especially for detailed and helpful feedback on chapter drafts.

I can't possibly thank my advisor, Alexander Williams, enough. Alexander's generosity as an academic mentor is truly grand (in at least the senses 1a, 1b, 6a, 6b, and 11a, as currently listed in the OED, and arguably in other senses as well). He has always been willing to meet and talk, or read an article together, or give feedback on a draft, or help organize a conference, or design a

series of psycholinguistic experiments. The existence of this document is in no small measure due to his dedication as a mentor and advisor. I cannot count the things I have learned from Alexander, or the ways in which this dissertation has been made possible and otherwise improved by way of his feedback and encouragement.

Thank you, Alexander, for your generosity, patience, guidance, and wit. It is not self-evident that a sandwich recommendation is insufficient to bridge the gap between what I can express in words and the gratitude I owe. So, I hereby recommend that you and Marcel should try the roast beef and the Italian at Victoria's in Hingham, Massachusetts (on the way out to the historic Nantasket Beach, which is in neighboring Hull), should the course of events ever bring you up that way.

To Louise Gilman—thank you for being a friend and also just the best person to work with. Many hard weeks were gotten through because of you.

I owe an immeasurable thanks to to my parents-in-law, Dan and Dody Williams, for love and encouragement and very generous support, without which I wouldn't have been able to get here. (Thanks also for being Emma's efficient causes.)

Thanks to all of my siblings, and to my abundant nephews. A special thanks in particular to my brother, Jim—for being a steadfast friend, always prepared with a wisecrack, or to re-watch an episode of *Rockford*, or to stare into the abyss together.

Many, many thanks to my father, Gerry, and my mother, Roberta. You both have been there from the start, in both the literal and the figurative senses of ‘been there.’ Thank you for your beneficence and support and realism and optimism and humor and so many other things—too many to count.

Above all, I’d like to thank my wife, Emma. Thank you, Em, for your friendship, patience, good humor, and tireless support over these many years. *A chuisle mo chroí*; Macushla; bean—a dozen new names each day would not suffice. I love you more than words can wield the matter.

Table of contents

Chapter 1: A cognitivist perspective and a Fodor-like hypothesis regarding the distinction between semantics and pragmatics

Dennis Eckersley, color commentator for television broadcasts of Boston Red Sox baseball games, often utters sentences like (1).[1]

(1) He's a pair of shoes.

When Eckersley (or "Eck") says this, he gets across to his regular audience the message that a certain player struck out without even swinging at the last pitch. In effect, Eck communicates that the player batted so poorly, an empty pair of shoes at the plate would have been equally useful. Sometimes, Eck gets across this message by uttering just 'Shoes!' Of course, (1) does not mean that a batter has struck out, and likewise the meaning of 'shoes' has nothing directly to do with baseball. The message about baseball is one that Eck means to communicate by uttering (1) (or by exclaiming 'Shoes!'), not what (1) itself means.

A distinction is commonly drawn between semantic and pragmatic aspects of linguistic understanding. Semantics is the study of the meaning of words (like 'shoes' and 'he') and how those meanings combine to yield the meanings of more complex linguistic objects, like the sentence (1). Meanings are characterized in semantic theory independently of considerations about the communicative intentions that speakers have on particular occasions of use. Facts about language use—to include facts about speaker meanings and how they are communicated—

1 (2021, May 11) *Boston Red Sox vs. Oakland Athletics* [television broadcast] New England Sports Network.

are the subject matter of pragmatics. Semantic and pragmatic facts somehow conspire to allow Eck to get his message across by uttering (1); accordingly, an explanation of this feat will appeal to both sorts of fact.

This dissertation asks what the semantics-pragmatics distinction comes down to for those who adopt a perspective known as 'mentalism.'[2] For mentalists, language is fundamentally psychological, and its study is therefore ultimately a branch of psychology (e.g., Chomsky 1965). In a mentalist linguistics, the question arises as to what psychological division grounds the semantics-pragmatics distinction, if any. Some mentalists defend a strong answer to this question, according to which the semantics-pragmatics distinction is grounded in the modularity of certain semantic processes (Sperber & Wilson 1986; Carston 2002b; Borg 2004; Harris 2020). The notion of modularity here is due to Jerry Fodor (1983). Fodor-modular processes have (among others) a feature called 'encapsulation.' Whereas non-modular, unencapsulated processes are potentially subject to the influence of any piece of information available to the agent and relevant to the task at hand, encapsulated processes (sensory processes like vision, paradigmatically) are blocked off from such influences and driven entirely by the bottom-up input or stimulus (together with information available to a restricted database, proprietary to the processes in question).

So, returning to (1): The encapsulation hypothesis for semantics has it that a modular semantic system decodes the linguistic input (the sentence (1)) itself, retrieving and combining meanings in a way that is not influenced by (for

2 The label appears to be originally due to critics of the perspective in question (e.g., Bloomfield 1933). The perspective is also sometimes called 'conceptualist' (e.g., Katz & Postal 1991).

example) the hearer's beliefs about the speaker's intentions or goals. Meanwhile, a non-modular pragmatic system draws inferences about what Eck intends to communicate, where such inferences are guided by information provided by the linguistic input, but also by information from various non-linguistic sources as well. In this way, the encapsulation hypothesis for semantics naturalizes the semantics-pragmatics distinction by locating it in a deep psychological division between two fundamentally different kinds of process. In what follows, I ask what processes might count as semantic, whether we have good reason to think they are modular in a Fodor-like way, and what the Fodor-modularity of these processes would entail regarding mentalistic semantic theory more generally.

I address these questions in light of available psycholinguistic methods and data, and in rapport with debates about encapsulated processing in other domains, including vision. The second and third questions (why think semantic processing is modular, and what would the modularity of such processes entail for linguistic theory?) are addressed in later chapters. In the present chapter, I articulate the hypothesis in more detail, situating it against its theoretical background, and specifying which functional components of utterance interpretation might be both encapsulated in a Fodor-like way and semantic in a narrowly linguistic sense of 'semantic.'

1.1 SEMANTICS VERSUS PRAGMATICS

Semantics, on one way of using the term, is the study of the meanings of words and the ways in which such meanings compose to yield the meanings of phrases.

There is disagreement about how to characterize word meanings, and disagreement about the principles of their composition. But most theorists are committed to some version of this idea: a semantic theory should pair words with meanings that can be characterized independently of considerations about particular occasions of word use. Speakers have communicative intentions and hearers often manage to infer them. But such considerations—about intentions, inferences, and shared knowledge—are not, strictly speaking, semantic considerations. That is, they are not considerations about words *per se* but considerations about their use. A pragmatic theory aims to explain such facts about how speakers and hearers use words. A semantic theory pairs words with meanings, characterized independently of the uses to which those words are put, and constructs phrasal meanings from the meanings of their constituent words, given certain principles of meaning composition.

When it comes to explaining various observations related to linguistic understanding, we sometimes face a challenge in discerning the respective explanatory contributions of semantic and pragmatic factors. This challenge arises because considerations that do not appear to concern the meanings of words—considerations about the intentions of a speaker, about background knowledge shared by speaker and hearer, etc.—often intrude in attributions of speaker meaning, even of the relatively direct sort. The point is familiar for communication that is quite clearly indirect. If you ask me to start a bonfire and I utter (2), to get across that we should start by burning the book, this aspect of my message cannot be read off the sentence alone.

(2) The book is old.

You will only understand this indirect message with reference to our goals, facts about the conversational context, and background knowledge. (Compare what is required to understand the message that Eck ultimately communicates with (1).)

But this is also true, if less flagrantly, for the more direct or "literal" aspect of my message. What decides whether, in uttering (2), I mean that the folio is old, rather than the story inscribed in it? At least at first blush, it is not the sentence itself that so decides, but again my intentions, which you can only divine in relation to our purposes, plus what we know about paper and fire. On its own, (2) does not decide what I mean, regardless of the relative directness of the speaker meaning in question. In uttering (2), I most likely intend to communicate either a message about a physical story vehicle, or a message about an abstract story content. And my audience will make a best guess as to my intention in this regard, in part based on their recognition that I have uttered (2) and in part based on their grasp of what (2) means, but also based on decidedly extra-linguistic considerations. The sentence (2) by itself specifies no single message. Hence, no feature of (2) itself specifies a single message. On the assumption that the linguistic meaning of (2) is a feature of (2), it follows that the linguistic meaning of (2) fails to specify (or 'underdetermines') a single message.[3]

It is uncontroversial that a given sentence, together with whatever its narrow linguistic meaning might be, underdetermines[4] (or fails to uniquely specify)

3 Cf. Sperber & Wilson 1986; Carston 2002b, 2012; Pietroski 2006, 2008, 2010, 2018.

4 'Underdeterminacy' is a clunky word but one that allows us to preserve 'indeterminacy' for a different purpose (cf. Carston 2002b: 20, from whom I borrow the terminology and the following rationale for deploying it). In saying that something is indeterminate, we generally mean that there

speaker meaning, since quite a lot of what speakers mean is conversationally implicated, and it is uncontroversial that sentence meaning underdetermines what is implicated (Grice 1961). The meaning of (2), whatever that turns out to be, does not specify the message that we should burn the book. However, not only does (2) underdetermine what is meant in this broad sense of 'what is meant,' on which implicatures are part of what is meant; (2) also underdetermines what is meant in a narrower sense.[5] This narrower sense of 'what is meant' can be characterized in different ways, often captured with terms of art like 'what is said,' or 'what is directly asserted.' The distinction between what is indirectly meant and what is directly meant is often illustrated with examples like (3) (Searle 1975).

(3) Student X: Let's go to the movies tonight.

Student Y: I have to study for an exam.

---

is no fact of the matter regarding it. So, for example, it might be proposed that the extension of a word is indeterminate, in the sense that there is no fact of the matter concerning what (for example) 'heap' is true of, or what counts as an instance of the action described by 'cut.' The underdeterminacy theses described here do not entail that what is meant is indeterminate. Underdeterminacy requires only that the facts of the matter concerning what is meant (whatever they are) are not determined by the sentence uttered and its narrowly linguistic meaning. That said, what is meant in the broad sense, to include implicatures, does seem to be indeterminate, at least in some cases. In uttering (2), for example, I need not have highly specific intentions regarding what I aim to implicate. By uttering (2), do I implicate that we should start the bonfire by burning the book, or that we should *try to* start the bonfire by burning the book, or that we *should at least consider trying to* start the fire by burning the book? Perhaps there is no matter of fact in this regard; if so, then what is meant in the broad sense is not only underdetermined by the meaning of the uttered sentence, but also indeterminate (cf. Grice 1975/1989: 40). Some have argued that what is meant in the more direct fashion is also indeterminate (see, e.g., Cappelen & Lepore's (2005) defense of 'speech act pluralism').

5 There is also a third underdeterminacy thesis, according to which what is meant in the narrower sense underdetermines what is meant in the broader sense. So, for example, what I directly assert (that the folio is old) underdetermines what I implicate (that it should be used to start the bonfire). This is also uncontroversial. As Carston (2002: 20) notes, the underdeterminacy thesis that is subject to substantive debate says that the sentence, together with whatever meaning it is assigned by a semantic theory, underdetermines what is meant in the narrower, direct sense.

By uttering the sentence she utters, Y has rejected X's proposal. But this rejection is a message that Y communicates indirectly, by way of communicating a more direct message about having to study, but also by way of reliance on X's knowledge about movies and exams. Compare the scenario described above for (2), where the speaker indirectly communicates a suggestion to use the book to start the bonfire by asserting that a certain folio is old and relying on the hearer's awareness of various facts about paper and fire. Again, the fact that pragmatic considerations intrude between recognition of the sentence and recovery of indirectly communicated messages is uncontroversial. But consideration of (2) shows that pragmatic considerations also intrude between recognition of the linguistic input and recovery of what is directly meant.

The sentence (2) underdetermines a proposition. Hence, whatever message a speaker directly communicates with (2), that message is not uniquely specified by (2) itself. And this underdetermination by (2) of a single message seems to be due to the word 'book,' which on its own fails to uniquely specify what I call a 'content.' As I use the term, a content is the sort of thing that might be a (proper or improper) part of something that is (more or less) truth-evaluable: concepts and thoughts; or referents, properties, and propositions; or truth-conditional contributions and truth conditions. These are the sorts of things with which semantic theories ordinarily pair words and sentences. However, consideration of words like 'book' gives rise to one of the salient lessons of recent philosophy of language, and one of the substantial challenges for theories of linguistic meaning: it is sometimes unclear whether a word—even a word that is ordinarily considered

so-called "content word," such as the noun 'book'—is in fact attached, as a matter of narrowly linguistic fact, to anything content-like.[6]

My loose use of 'content' here glosses over certain points of disagreement. In what follows, I take a translational approach to semantics. A translational semantics gives the meaning of natural language expressions by, in effect, translating them into another system of representations—typically, expressions in a language of thought, or concepts and combinations thereof. So, on a translational approach, a word like 'cat' might be paired with a mental symbol, CAT, an atomic concept that allows us to represent cats as cats. On this line of thought, to give the meaning of 'cat' is just to specify its translation in the language of thought (see, e.g., Fodor 1989).

Some (e.g., Lewis 1970) argue that a translational semantics is not, properly speaking, a semantic theory. Some theorists insist that a semantic theory must (by definition of 'semantic') specify representational relations for which one relatum is a thing out there in the world, a thing that gets represented without representing anything else. A proper semantic theory, on this view, cannot pair natural language expressions with expressions in another language.

Since a translational semantics does not pair natural language expressions with things in the world, it is sometimes—incorrectly, as I shall presently argue—criticized as vacuous. The complaint against the translational approach to

---

6 See citations above, fn. 3. This lesson is one that emerges from Wittgenstein's (1953) reflections on the many ways a single word is used. But Wittgenstein's point is also prefigured in many passages of Aristotle (e.g., *Categories* 15b16-31; *Physics* I.7 190a31; *De Anima* 412b5; *Metaphysics* IV.2; *Nicomachean Ethics* 1.6 1096a23). Following up on Wittgenstein, the lesson is then reinforced by other ordinary language philosophers, including Austin (1961, 1962). The idea that words fail to determine anything content-like is also central to internalist conceptions of linguistic meaning (e.g., Chomsky 1964), as discussed in what follows (cf. Pietroski 2017).

semantics is something like this: “I asked what the word ‘cat’ means, and you’ve told me that ‘cat’ means the concept CAT, and now I want to ask what the concept CAT means. Is the deferral going to continue?”

This complaint about translational semantics is fair enough, but ultimately misguided. If you wanted your semantic theory to link words with the world, then a translational semantics that pairs words with concepts does involve deferral of the very question you were intending to ask when you asked what ‘cat’ means. Furthermore, at least some of my uses of ‘cat’ do seem to correlate fairly well with the presence of actual cats, or pictures of cats, or etc. There must be *some* relation between the word ‘cat’ and cats, it seems. If there is such an explanation, then it is not unreasonable to expect it to be discovered. And we could easily call a theory that explains the relation between ‘cat’ and the cats a ‘semantic’ theory.

A common way to pursue a non-translational semantics is by way of the notions of truth and reference. To give the meaning of a sentence is to give an account of the conditions under which it is true; and to give the meaning of a word is to give an account of the contribution it makes to the truth conditions of sentences that contain it. A well-known statement of this view is due to David Lewis (1970: 18ff.). But even Fodor, who defends a translational approach to linguistic interpretation, agrees that, strictly speaking, a semantic theory is pitched at the level of thought, since it is the concepts, not the words, that bear relations of truth and satisfaction to things in the mind-external world (1989: 418ff.).

In any case, the fact that a translational semantics does not directly pair natural language expressions with mind-external entities does not render such a

theory vacuous (Fodor 1989: 418ff.). By specifying the mind-external entities that our concepts represent and the truth-conditions of the thoughts that are built out of such concepts, we can give a "real" semantics (of the sort that Lewis calls for), albeit one that is pitched not at the level of natural language but instead at the level of thought. We might then call this "real" semantics a 'truth-conditional semantics for expressions in the language of thought.' We can additionally specify translations between natural language expressions and expressions in the language of thought and call this a 'translational semantics for natural language expressions.' And the latter theory is not necessarily vacuous, nor is it unworthy of, or otherwise unfit for, the label 'semantic.' Insofar as the expressions of natural language derive their contents from the language of thought representations to which they are attached via a translational semantics, the translational semantic theory escapes the charge of vacuity (Fodor *o.c.*; cf. Carston 2002b: 56ff.).

The phenomenon of polysemy, as exhibited by 'book,' raises a challenge for any semantic theory for natural language expressions, translational or truth-conditional. Polysemy is importantly distinct from homophony. Two words are homophones just in case they are distinct words that happen to share a pronunciation. For example, I can utter the sound of 'the bank is old' to communicate a either a financial or a fluvial message. And this is because /bæŋk/ is the sound of two distinct words, one of which has to do with finances and the other with rivers. We can distinguish these two words with numerical subscripts, such that '$bank_1$' is the word having to do with (say) finances and '$bank_2$' is the

word having to do with waterways. Since there are two words, '$bank_1$' and '$bank_2$,' there are in fact two different sentences corresponding to the string 'the bank is old.'

(4) The $bank_1$ is old.

(5) The $bank_2$ is old.

Homophony, then, is not a case in which the meaning of the word underdetermines a content; rather, it is a case in which a word-form underdetermines an index, and so (in effect) underdetermines word identity. If we assume that semantic processes operate over words and that homophony resolution involves identification of the word being tokened, homophony resolution is 'pre-semantic.'[7] Homophony is therefore compatible with the idea that a semantic theory pairs each word with a single content, since its resolution requires pragmatic intrusion only between the identification of the phonological word-form and an index (or identification of the tokened word or root), not between word recognition and content determination.

Polysemy resolution, by contrast, does raise a substantial challenge for semantic theories that aim to pair words with contents. Again, suppose '$bank_1$' is the word that has to do with finances. Even given a resolution of this homophony, (4) exhibits a further indeterminacy. Just as I can use 'book' and therefore (2) to communicate a message either about a folio or a story, likewise I can use '$bank_1$' and therefore (4) to communicate either a message about a socially constructed financial institution, or one about a particular building. Which of these messages

7 In fact, from the processing-oriented perspective that I take on such questions, it becomes hard to insist that homophony resolution is entirely pre-semantic. See further discussion in section 1.5.

is paired with (4) by a semantic theory? What content is paired, as a matter of linguistic fact, with '$bank_1$?' One might seek to respond here with more diacritics. However, as I show in what follows (chapter 4), there are good reasons to preserve the distinction between homophony and polysemy, and good reasons to expect that their resolution is effected via different mechanisms.

Again, what content is paired, as a matter of linguistic fact, with '$bank_1$?' Here, one might compare the polysemy of 'book' and '$bank_1$' with the context-dependence of the reference of an indexical such as 'this.' A semantic theory cannot pair 'this' with any of the things that we refer to in using it. As Strawson notes, "if someone asks me the meaning of the expression 'this' … I do not hand him the object I have just used the expression to refer to, adding at the same time that the meaning of the word changes every time it is used. Nor do I hand him all the objects it ever has been, or might be, used to refer to" (Strawson 1950: 328). Instead, to characterize the meaning of 'this,' I appeal to something that is abstract with respect to anything content-like (referent, concept, truth-conditional contribution). A common strategy in formal semantics is to pair words like 'this' not with a content but with what is called a 'character,' a function from contexts to contents. So, while (6) does not on its own determine a proposition, it does so relative to a contextual assignment of values to variables in logical form (viz., assignment of a referent to 'this').

(6) This is old.

This, in any case, is the standard approach to indexicals in formal semantics (e.g., Kaplan 1989). Words like '$bank_1$' and 'book' do not appear to be indexicals,

however. At least at first blush, polysemy and indexicality are different categories of underdeterminacy. In what follows, I rehearse and reinforce arguments for the conclusion that this difference (between polysemy and indexicality) is also an important one (section 4.3.2).

Resolving the polysemy of 'book' does not seem to be a case of inferring a message that is communicated by virtue of the fact that the speaker more directly communicated another message, more closely tied to the sentence uttered. There is no single message that is closely tied to the sentence itself. Instead, pairing linguistic input that includes 'book' with any message whatsoever seems to require appeal to pragmatic considerations. What, after all, is the literal or minimal sense of 'book'? What single concept is attached to 'book,' given that we can use the word to communicate a message about a physical folio, or instead one about an abstract story? Is it FOLIO or STORY that a semantic theory should pair with 'book'? More or less the same question arises for the truth-conditional semanticist. What is the truth-conditional contribution of 'book,' a clause about folios, or one about stories? And we can also translate the puzzle into the idiom of reference. What is included in the extension of 'book,' the folios or the stories? The word itself decides none of these questions, each of which is settled only relative to speaker intentions, and in a way that cannot be assimilated to homophony, indexicality, or figurative use.

Two important lessons emerge from consideration of polysemy. First, there are disambiguated, non-indexical words that apparently do not specify a single content. Second, it follows that there are some sentences that do not specify a

single message, not even a literal or highly general message of the sort that semantic minimalists favor (see, e.g., Borg 2004; Cappelen & Lepore 2005; see also sections 1.3.2.2, 4.1.1, and 4.3.3), and this underdetermination is not due to indexicality. In light of these two lessons, semantic theorists have two options, broadly speaking. The first option is to keep semantic theory in the business of pairing words with contents and deriving from these (and rules of composition) the contents of phrases. However, to do so, we must allow an increased role for pragmatic considerations in semantic theorizing. One such option would be to pair 'book' with a list of contents, and then to incorporate in our semantic theory a context-sensitive compositional rule that selects one of these and combines it with a context-sensitively selected sense of 'old' to yield the sense of 'old book' (cf. Pustejovsky 1995).

Alternatively, we can exclude such pragmatic considerations from semantic theory by pairing a word with something that is abstract with respect to anything content-like. If one takes Strawson's lesson about 'this' seriously and then attends to 'book,' it is plausible to conclude that, at best, a semantic theory can pair words with something like a general instruction for use. 'Book' is not paired with either of the concepts that a speaker might use it to express (FOLIO, STORY). Instead, 'book' is paired with something that is abstract with respect to anything content-like—something neutral with respect to the specific concepts the word can be used to express (and likewise with respect to the entities it denotes, or the truth-conditional contributions that it makes). It is not easy to characterize what such abstract word meanings might be without deploying a metaphor. Strawson

appeals to “general directions for use in making true or false assertions” (Strawson 1950: 327). There are other apt metaphors, each of which seems to capture a slightly different model of word meaning: a word like ‘book’ might be paired with something like a constraint on expressible contents (Harris 2020), or an instruction to fetch one concept from an address that hosts others (Pietroski 2018), or a pointer to a vaguely defined region in conceptual space (Carston 2012). (See further discussion of these options in chapters 4 and 5.)

Just about everyone agrees that some admixture of semantic and pragmatic factors is needed in an explanation of linguistic understanding, even when we restrict our attention to what is meant by a speaker in the relatively direct fashion. At issue are the correct proportions: Which part of the broader theory of linguistic understanding should do which job? In particular, what should a semantic theory pair with a word like ‘book,’ such that one and the same word can be used to directly communicate a message about an abstract story content or a message about a concrete story vehicle? These are open questions. However, one thing that most parties apparently agree on is that a semantic theory ought to be in the business of pairing a word with *something or other*—if not a single content, then a list of contents, or perhaps a rough indication of a space of contents. This commitment ends up connecting the semantics-pragmatics distinction with another distinction altogether.

1.2 FODOR-MODULARITY

In cognitive science, a distinction is often drawn between processes that are modular and processes that are not modular. In this section, I characterize one way of thinking about modularity, due to Jerry Fodor's (1983) *Modularity of mind*. I describe the features that Fodor attributes to modular processes and illustrate the distinction between modular and non-modular (or 'central') processes. I emphasize in particular the importance of the notion of encapsulation in Fodor's account of modularity. Modular processes are encapsulated, in Fodor's view, in the sense that their operations are driven by the bottom-up input and blocked off from the top-down, extra-modular influence of most of the information available in the mind more generally. The sensory processes, including vision, are paradigmatic examples of Fodor-modular processing systems.[8] The idea to be pursued in what follows is that semantics should be added to the list of Fodor-modular processing systems.

Fodor characterized a three-part functional taxonomy for cognition: the mind consists of 'transducers' that reformat proximal stimulations, a central processing system in which beliefs are fixed, and input systems that mediate between transducers and central processing (1983: 41). According to Fodor, input systems are a natural kind—a class of phenomena that have interesting scientific properties over and above those that define the class (1983: 46). Input systems are the cognitive systems that count as modular in Fodor's sense of the term. Fodor identified nine properties that modular processes have in common and in contrast

[8] However, there is ongoing debate about whether even sensory processes are Fodor-modular, as I briefly discuss here (and in more detail in chapter 2).

with non-modular or central processes: they are domain-specific; their operation is mandatory; they are fast; central processing has limited access to the representations they compute; they are informationally encapsulated; they have representationally 'shallow' outputs; they are associated with a fixed neural architecture; they exhibit characteristic and specific breakdown patterns; and their development exhibits a characteristic pace and sequencing (1983: 47-101).

As discussed in more detail below, Fodor's is not the only conception of modularity. Fodor's (1983) is a performance-oriented notion of modularity, on which it is a feature of certain processes or processing systems; this processing-oriented conception of modularity is my focus in most of what follows. However, there is also an alternative, competence-oriented notion of modularity, one that is articulated by Chomsky (1971, 1986), on which a module is a body of information (perhaps innately specified) that underlies some capacity. Importantly, the operations that underlie the actual use of the capacity may or may not be restricted to the influence of the proprietary information in a Chomsky-module. In other words, the two notions of modularity do not stand or fall together. However, we can also think of a module as a hybrid of a Fodor- and a Chomsky-module, such that a module is a restricted body of information plus some processes that operate over only that information. This three-way distinction among different notions of modularity is characterized by Emma Borg (2004: 75-76) and discussed in more detail in what follows. For now, however, my attention is restricted to the notion of a Fodor-module, a process or processing system that exhibits the nine features mentioned above.

Fodor's own list of modular processing systems is relatively short, including sensory processing systems (for vision, audition, etc.) and a handful of linguistic processes that are similarly perceptual (1983: 44). Fodor occasionally suggests a modest expansion of the list, perhaps to include mindreading or "Theory of Mind" among the Fodor-modular processes (Fodor 2000: 97). In any case, the Fodorian hypothesis posits a relatively small set of encapsulated input systems that provide representations of the distal environment to central cognition, supplying an evidentiary basis for belief fixation, where the latter results from unencapsulated or isotropic inferential processes, susceptible to the influence of any piece of information available to the mind.

Here is how Fodor would contrast visual processing with the process of making a decision, for example. Visual processing—processing of the signals coming from the retina—seems to be domain-specific, in that its operations are restricted to a certain kind of input—signals from the retina. In contrast, any kind of information might play a role in the process of decision-making, which is not limited to certain special kinds of input. Additionally, we have minimal executive control over visual processing. Provided that my eyes are open, I see a visual array as consisting of objects distributed in three-dimensional space, whether I want to or not (1983: 53). Decision-making, by contrast, is a process over which I have considerably more executive control. Visual processing is also generally much faster than decision-making. Its relatively fast pace is in part due to the fact that visual processing exhibits a kind of two-way informational isolation. The operations of the visual system are inaccessible to conscious introspection.

Decision-making, by contrast, is a process regarding which we have much more in the way of introspective access. The second half of the two-way informational isolation exhibited by modules is what Fodor called 'encapsulation': most of the information that is available to central cognition is not able to influence the operations of the visual system. Whereas any relevant piece of information might play a role in decision making, quite a lot of relevant information seems to be incapable (in principle, not just in practice) of influencing visual processing.

For Fodor, encapsulation is an especially crucial feature of modular systems: the flow of information is always from transducer to input system and then to central cognition, in exactly that order. That Fodor took encapsulation to be an especially important feature of modules is evidenced by the fact that he spends about twenty-two pages in his 1983 motivating the claim that modular systems are encapsulated and defending this claim against apparent counterevidence (1983: 64-86); by contrast, only about four pages are spent defending the claim that the operations of modular systems are relatively fast (1983: 61-64). This disproportionate attention is presumably because the two-way isolation exhibited by modules is a constitutive or necessary feature of modules, whereas a fast pace of processing is not constitutive of modularity. To be a Fodor-module is to be the kind of system that passes information only in one direction. Even if modular processes are all relatively fast-paced, it is not their speed that makes such processes Fodor-modular. Arguably, what makes a process Fodor-modular is two-way informational isolation, of which encapsulation is one half (along with

inaccessibility). Modular processes might happen to be fast paced, but they must be encapsulated.

The notion of encapsulation that Fodor takes to be constitutive of modularity is a strict one. Fodor-modular processes are driven entirely by two things: the stimulus or bottom-up input, and whatever principles and information are available within the modular system itself. Information that is not made available as part of the input or the system's proprietary database is incapable, as a matter of architectural fact, of influencing the internal operations of a Fodor-modular processing system. It is not just that, during the moment-by-moment processing of visual inputs, the visual processing system happens to be insusceptible to the top-down influences of (say) my desire for a sandwich. My propositional attitudes, to include my beliefs, goals, and desires, are *incapable in principle* of having any influence on early visual processing. As understood by Fodor, encapsulation is incompatible with any such top-down influence of extra-modular information on the internal operations of a modular system. This Fodorian architecture is attractive in part because, given this strict understanding of encapsulation, the notion of modularity promises to cut psychological nature at a real joint (cf. Firestone & Scholl 2016: 1).

That said, there is ongoing debate concerning the extent to which (for example) visual processing is encapsulated.[9] A large body of experimental work

9 For a sustained defense of hypothesis that visual processing is encapsulated in Fodor's sense, see not only Fodor 1983 but also Pylyshyn 1999. A more recent defense of the claim that sensory processes in general are Fodor-modular is given in Firestone & Scholl 2016. For an overview of the criticisms of Fodor's view of modularity, and in particular his conception of encapsulation, see Carruthers 2006. For a sustained argument against a Fodor-like conception of encapsulation for visual processing, see Ogilvie & Carruthers 2016.

suggests that sensory processes are susceptible to top-down influences of (for example) the beliefs, desires, and goals of the agent. If this interpretation of the relevant results is correct, then even sensory processes like vision are not encapsulated, at least not in Fodor's strict sense. As just one example, some studies indicate that objects appear to be closer when they are desired by the agent (e.g., Balcetis & Dunning 2010). On one interpretation, this observation is taken to show that visual processes are susceptible to top-down influences of extra-modular information. Information about what I desire is not proprietary to the visual processing system, but my desires appear to modulate the manner of visual processing.

This, in any case, is one interpretation of the relevant observations. Other interpretations are compatible with Fodor's conception of encapsulation. This invites further experimentation and more discussions about how to interpret the relevant results; and the experimental literature in this regard is substantial. The question becomes an abductive one: What is the best explanation for the available experimental observations concerning perceptual (and other kinds of) processing: one that appeals to Fodor-style encapsulation or one that appeals to a weaker notion of encapsulation? I return to this question in chapter 2.

1.3 On the relation between the two distinctions

Fodor's distinction between modular and non-modular or central processes is meant to ground the distinction between perception and cognition. The distinction between semantics and pragmatics is a distinction between two aspects of

linguistic understanding. In what way do the two distinctions relate to one another? First, there are respects in which the dialectics that arise concerning the two different distinctions parallel each another. The way in which the debates about encapsulation have played out is quite like the way in which the debates about separating semantic theory from pragmatic considerations have played out. However, related (but in addition) to the dialectical similarity, there is perhaps also a deeper connection between the two distinctions. It has been suggested that the semantics-pragmatics distinction is, in effect, a species of the more general distinction between modular and central processing. If the semantics-pragmatics distinction can be grounded in the modularity of semantic processes, then we might conclude that this theoretical distinction indeed cuts psychological nature at a real joint. And there are some arguments worth considering on behalf of this line of thought.

The encapsulation hypothesis for semantics, as I have so far characterized it, claims that certain components of utterance interpretation are both encapsulated and, in some sense, semantic. It remains to be said what these functional components of utterance interpretation are and in what sense they are semantic. I discuss two such functions in what follows: assignment of indices that constrain the contents with which a word-form might be paired, and assignment of contents to indexed word-forms. The encapsulation hypothesis for semantics might be restricted to index assignment but might also be taken to include content assignment. As I discuss in the present section, it is unclear which (if either) of these two hypotheses Fodor himself endorses. Fodor for the most part attends to

those linguistic processes that are perception-like, including syntactic and phonological processes. However, there is some reason to think that Fodor additionally endorsed at least one of the two versions of the encapsulation hypothesis for semantics that I have characterized here—that is, either a version according to which index assignment is encapsulated, or one according to which content assignment is encapsulated.

However one settles the exegetical question regarding Fodor's own view, several other theorists have defended the Fodor-like hypothesis according to which content resolution is the result of an encapsulated process, including both semantic minimalists (Borg 2004) and Relevance Theorists (Sperber & Wilson 1986. The attraction of this general line of thought will depend on one's views about the subject matter and goals of linguistic theory. I therefore characterize a certain framework for linguistic theorizing about semantics and pragmatics, a perspective that I call 'cognitivist,' adapting a label used by some Relevance Theorists. As I characterize the perspective, cognitivism consists of a set of theses or commitments, including a mentalist view of linguistic ontology, a commitment to an internalist and translational semantics, and a focus on processing and the perspective of the hearer. It is within this cognitivist framework that the connection between semantics and modularity appears to be especially close.

1.3.1 PARALLEL DIALECTICS

There are parallels between debates about encapsulation, on the one hand, and debates about how to characterize linguistic meaning independently of pragmatic

considerations, on the other hand. In this section, I briefly point to these parallels, which are then pursued in more detail throughout the dissertation. As I go on to immediately discuss in the subsequent section, there is also a potentially deeper and more interesting connection between linguistic meaning and modularity. But the dialectical parallels, while familiar, are worth pointing out. They set the stage for discussion of the more interesting connections.

Here is a caricature of the first dialectic, concerning modular versus non-modular processing—and, in particular, concerning encapsulation as Fodor understands it. Whatever is involved in sensory processing, its outputs are then deployed in the drawing of various inferences about the distal causes of the sensations, inferences that result in fixation of beliefs about the world outside our heads. Inference and belief fixation are subject to the influence of whatever relevant information might be available to us, not just the information provided by processing of sensory input. However, according to the Fodorian, the sensory processes themselves (along with a few other perception-like processes) are not subject to such top-down, extra-modular influences. The operations of the visual system itself are encapsulated, being driven entirely by the bottom-up input, the principles that govern the operations of the visual system, and whatever information is stored as part of that system's proprietary informational database. There is *something*—a geometrical sketch of the visual scene—that is uniquely determined by the input, something that can be characterized independently of pragmatic considerations about the agent's beliefs, desires, goals, etc. Others doubt the Fodorian's claim, arguing that there is a lot of evidence for top-down,

extra-modular influences on even early stages of visual (and other sensory) processes. Perhaps, after all, we should not insist that modular processing is encapsulated from top-down pressures in the way Fodor hypothesized. Aware of this challenge, some proponents of Fodor's notion of modularity defend their view against the apparent counterevidence. Others articulate alternatives to Fodor's conception of encapsulation and modular processing. The debate is ongoing.

Here is a caricature of the second dialectic. The linguistic meanings of the sentences we encounter give rise to various inferences, inferences that result in beliefs about what speakers mean, and that therefore require appeal to an (in principle) unlimited number of potentially relevant pieces of information that are not provided by the linguistic input itself. Still, the semanticist usually maintains that there is *something* meaningful (a highly general concept or proposition, a propositional schema, an instruction to fetch and combine concepts—*something*) that is uniquely determined by the linguistic input, something that can be characterized independently of speaker intentions or hearer inferences. Others are skeptical, noting that there is a lot of evidence for pragmatic intrusion in just about any effort to characterize what words (including 'book') and sentences themselves mean. Perhaps, after all, we should not insist on being able to characterize word meanings independently of such pragmatic considerations. Aware of this challenge, many semanticists devise ways to stay in the business of characterizing word and phrase meanings independently of pragmatic

considerations. Others propose new approaches that blur the lines between semantic and pragmatic processing. The debate is ongoing.

Arguably, these dialectical parallels arise in part because appeals to narrowly linguistic or semantic meaning, like appeals to encapsulation, are in part meant to explain how our minds solve 'frame problems' (Fodor 1983). Any fact or belief might be relevant to confirmation or disconfirmation of any given hypothesis (Duhem 1906/1964, Quine 1960). We cannot non-arbitrarily exclude in advance of inquiry or interpretation any fact or belief from the set of potentially relevant pieces of information. How, then, do we ever manage to fix a belief in a finite amount of time? In short, encapsulation is meant to solve such a frame problem—and so is the notion of narrowly linguistic meaning. Fixing beliefs about the world would be subject to an insurmountable frame problem if it were not for the operations of encapsulated input systems, which make available (to central cognition, where beliefs are fixed) representations that result from deterministic processes that are insusceptible to the influence of most of the information available in the mind in general. In other words, the encapsulation of certain input systems serves to constrain the central process of belief fixation, solving a frame problem.

Regarding utterance interpretation, a similar frame problem arises. The goal of utterance interpretation is to fix a belief about a speaker's communicative intentions—to figure out what messages the speaker intends to communicate. Recall, for example, Eck's utterance of (1). In principle, any piece of information whatsoever might be relevant to this task. How do hearers manage, in a finite and

often quite short amount of time, to fix beliefs about communicative intentions? The basic idea is that the linguistic input serves to constrain the inferential process of utterance interpretation by constraining the space of available hypotheses. In part, the claim that linguistic input carries, on its own, a meaning of one sort or another is meant to contribute to an explanation of how the frame problem is solved for utterance interpretation. So, for example, the fact that 'shoes' and (1) mean what they mean somehow constrains the available hypotheses about what Eck intends to communicate in uttering those expressions. This, in turn, recommends that we consider the possibility that linguistic meaning is recognized by way of an encapsulated, Fodor-modular processing system.

1.3.2 THE ENCAPSULATION HYPOTHESIS FOR SEMANTICS

On a common way of relating the two distinctions, semantic processes are modular and hence encapsulated, while pragmatic processes are non-modular and unencapsulated. In this regard, there is agreement among theorists who otherwise diverge quite a lot in their views about linguistic understanding, including certain semantic minimalists (Borg 2004) and Relevance Theorists (Sperber & Wilson 1986).

Before discussing these accounts, I first discuss in the immediately subsequent section whether Fodor himself endorses any version of the encapsulation hypothesis for semantics. As I show, there is some reason to think that Fodor would agree with a version of the encapsulation hypothesis for semantics, on which an encapsulated system carries out translation of natural

language expressions as input to expressions in a language of thought as output. I note that, for Fodor, such translations are not semantic in the Lewisian sense. In short, for Fodor, to give a translational semantics for a natural language is not to give a truth-conditional semantics. The latter can only be stated at the level of thought, not language; hence, Fodor's claim that a "real," truth-conditional semantics is not a level of linguistic description (1989: 418ff.). This, however, is perfectly compatible with the hypothesis that an encapsulated processing system maps natural language inputs to expressions in the language of thought (cf. Borg 2004: 83ff.).

1.3.2.1 FODOR EXEGESIS

Fodor notes that a standard taxonomy of mental processes would group together the sensory processes (vision, audition, olfaction, etc.) separately from representational processes (thought and language). However, as Fodor argues, "for the purposes of assessing the issue about modularity, a rather different taxonomy proves illuminating," one that identifies as modular "the perceptual processes *plus language*" (Fodor 1983: 44; original emphasis). What do perceptual and linguistic processes have in common, such that they form a natural kind, in Fodor's view? For one thing, utterances must be perceptually identified: you have to recognize a certain auditory stimulus as an instance of a phoneme, or word, or phrase—at least, if you care to make sense of the speaker's action.

As Fodor discusses, the perceptual identification of such linguistic entities (morphemes, words, phrases) presumably involves structural descriptions of the

kind involved in computing a token-to-type relation (recognition of an instance of a certain morpheme, etc.), the very sort of operation that input systems typically perform (Fodor 1983: 44). More importantly, "just as patterns of visual energy arriving at the retina are correlated, in a complicated but regular way, with certain properties of distal layouts, so too are patterns of auditory energy that excite the tympanic membrane in speech exchanges" (1983: 44). In addition, sensory and linguistic processes "both serve to get information about the world into a format appropriate for access by such central processes as mediate the fixation of belief" (1983: 46).

So, according to Fodor, some linguistic processes are modular—namely, those that are akin to perceptual processes in the aforementioned respects. The process of utterance interpretation as a whole is a matter of belief fixation. A hearer does not perceive what a speaker means; hearers must fix a belief (or beliefs) about speaker intentions, and they must do so by way of inference. Hence, utterance interpretation as whole is a non-modular or central process, though some of its constitutive processes may be modular (see Allott, forthcoming, for a detailed argument in support of this conclusion). This allows that certain components of utterance interpretation might be Fodor-modular—namely, those that appear to be more perception-like and less inferential. For example, Fodor himself (1983) argues that phonological processing and syntactic parsing are modular in his sense of the term, and hence encapsulated from the influences of general cognition. Phonological and syntactic processing do seem to have some characteristics of the paradigmatically perceptual processes: they involve

recognition of a stimulus as being a token of a certain type, and the results of recognition are then fed to the inferential processes involved in fixation of belief (for example, about speaker meaning).

Are any linguistic processes encapsulated in Fodor's view, but also semantic? There is some reason to think so, at least if we restrict our attention to a translational conception of semantics. Regarding the fact that modular operations are in some sense mandatory, Fodor notes as an example the fact that "you can't hear speech as noise *even if you would prefer to*" (1983: 53; original emphasis). From this, we are meant to conclude, at a minimum, that one cannot help but hear speech as consisting of phonemes and as having a certain hierarchical structure, and hence that phonological and syntactic processing are modular. Fodor does not explicitly defend the additional claim that something like grasping sentence meaning is mandatory. However, he does gesture in the direction of such a claim when he writes, "'I couldn't help hearing what you said' is one of those clichés which, often enough, expresses a literal truth; and it is what is *said* that one can't help hearing, not just what is *uttered*" (1983: 55; original emphasis). We are meant to conclude, it would seem, that grasping what is said is a mandatory and hence modular process. To grasp what is said, however, is to do more than to recognize phonological and syntactic structure. To grasp what is said, I need to do more than recognize the sentence you have uttered; to grasp what is said is to grasp something meaningful, content-like, and perhaps truth-evaluable. Granted, strictly speaking, Fodor claims here only that recovery of what is said is in some sense mandatory. However, given the context of this claim, in which he is giving

examples of processes that exhibit the features that are jointly sufficient for modularity, we are led to conclude that the process of recovering what is said (a content) is, in Fodor's own view, modular and hence encapsulated. This would seem to be a version of the encapsulation hypothesis for semantics.

Other passages appear to reinforce this reading. Fodor discusses sentence comprehension to illustrate the claim that modular processes are inaccessible to conscious awareness. He notes that "sentence comprehension, for example, involves not only acoustic encoding but also the recovery of phonetic and lexical content and syntactic form," where the operations at each of these stages are inaccessible to introspection (1983: 56). Again, we are led to conclude that these components of sentence comprehension are modular, to include recovery of lexical content. What does recovery of lexical content involve? On one reading, Fodor might be taken to suggest that recovery of conceptual content that is paired with a lexical item is an encapsulated process—again, a version of the encapsulation hypothesis for semantics. So, there is some reason to believe that Fodor indeed may have endorsed a version of the modularity hypothesis for semantics.

However, it is possible that Fodor means by 'recovery of lexical content' only the recognition of words. Perhaps the semantics-pragmatics distinction is ultimately grounded in the fact that word recognition (in effect, index assignment, rather than content assignment) is Fodor-modular. I discuss this possibility in what follows (section 1.5). First, though, a qualification is in order. According to Fodor, "understanding natural language expressions [is] a matter of *translating*

them" into the formulas of a language of thought (what Fodor sometimes calls 'Mentalese;' e.g., Fodor 1989: 418ff.). Monomorphemic words (for example, 'dog') have their meanings or contents by virtue of being connected to atomic concepts (for example, DOG). Such atomic concepts, the primitive elements of the language of thought, have their meanings by virtue of asymmetric dependency relations to entities in the mind-external world. Hence, on Fodor's view, there is no substantial lexical semantics; monomorphemic words are paired with atomic concepts, in such a way that the content of one monomorphemic word is independent of the content of any other. Hence, there is no "lexical decomposition;" what compositionality there is in natural language is inherited from the language of thought. (See, e.g., Fodor & Lepore 1998 for a defense of this overall picture of the relation between natural language and the language of thought.)

Again, for Fodor, "to understand a sentence of English … *just is* to compute its M[entalese]-translation" (1989: 418). Likewise, to understand a word of English is to compute its language of thought translation, which consists merely in recovering the atomic concept to which the word is attached. Is computation of the language of thought translations of natural language expressions an encapsulated process, in Fodor's sense? Again, this seems to be the suggestion of the two quotes from Fodor 1983 discussed above, concerning recovery of 'what is said' and lexical content. However, an important qualification is needed. For Fodor, computation of language of thought translations of natural language expressions is a syntactically driven process—driven, that is, by formal features

of natural language expressions, not by considerations about their content. That language of thought translation is computational in this way supports the idea that it is carried out by a Fodor-modular and encapsulated process. However, the fact that computation of language of though translations is syntactically driven also entails that a theory of content lies outside the scope of linguistic theory, as the latter is strictly understood. Hence, as Fodor puts it, while "it is very widely assumed, among cognitive scientists at least, that semantics is *a level of linguistic description*, just like syntax or phonology … in fact, this is all wrong" (Fodor 1989: 419; original emphasis). Here, Fodor's claim is only that translation from natural language to the language of thought is driven by formal, syntactic properties, and not by properties that are semantic in the sense that they involve relations to mind-external phenomena. A description of contents is thus beyond the scope of a theory of natural language expressions, being instead the purview of a theory of the language of thought.

This, however, leaves open the possibility that a modular processing system effects the translation between English and the language of thought during the moment-by-moment processing of linguistic stimuli. The system in question would take as input (mental representations of) expressions of a natural language and yield as output expressions of the language of thought. If the system is informationally encapsulated, then its operations are insusceptible to the top-down influences of general cognition. Given a particular word as input, the system in question would fetch the concept that is its uniquely determined translation in the language of thought, where this fetching operation is insusceptible to top-down,

extra-modular effects (for example, of inferences about speaker intentions). On a naïve version of the proposal, given ‘cat’ as input, the system context-invariantly fetches CAT. And given ‘orange cat’ as input, the system composes the concepts encoded by the constituent words, where the manner of composition remains constant across contexts of use. This is a version of the encapsulation hypothesis for semantics, on which the task of an encapsulated semantic processor is to map words to atomic concepts and phrases to combinations of atomic concepts.

In contrast with his discussion of phonological and syntactic processes, Fodor does not explicitly defend the claim that translation of monomorphemic words in a natural language to atomic concepts in the language of thought is the result of an encapsulated process. This is presumably because consideration of ‘cat’ and CAT suggests that translation of monomorphemic words, given a resolution of any homophony, is one-to-one. If there is no decision to be made about how to translate ‘cat’ into the language of thought, then there is no decision that might be subject to top-down influences of general cognition. When the linguistic processing system recognizes the word ‘cat,’ the concept CAT comes along for free. Hence, there is no work to be done in defending the claim that a certain amount of semantic processing is encapsulated—namely, retrieving CAT given recognition of ‘cat.’ Granted, it could be that the atomic concepts encoded by words are also attached in long-term memory to “files” consisting of various pieces of encyclopedic information about the things that satisfy the concept (cf. Sperber & Wilson 1986). And which pieces of information in the associated file are accessed along with the atomic concept might be subject to top-down effects

of general cognition. However, this is compatible with language of thought translation itself involving a kind of encapsulated content resolution that consists only in the fetching of the atomic concepts encoded by monomorphemic words and the combination of those concepts in accord with the structure of morphologically or syntactically complex input.

And, so understood, content resolution might be the result of a processing system that is encapsulated in Fodor's sense. Even if Fodor did not himself defend this hypothesis, it is plausible to think that he would have been attracted to it, especially given the quotes considered above (1983: 55-56). In any case, whether or not I am correct that Fodor may have found this view attractive, it is one that others have defended, as I discuss in the subsequent section.

1.3.2.2 Proponents of Fodor-modular semantics

A Fodor-like hypothesis about content-resolution is defended by semantic minimalists (Borg 2004) and by Relevance Theorists (Sperber & Wilson 1986). I describe each of these two accounts in the two subsections to follow. Both have much in common with what Fodor says about "M-translation," or translation between natural language expressions and expressions in a language of thought. There are important points of disagreement, as I note—between Fodor and Borg regarding what counts as a level of linguistic description, and between Borg and the Relevance Theorists concerning whether semantic processing, given a complete sentence as input, necessarily eventuates in a truth-evaluable proposition. However, concerning narrowly linguistic understanding, as separated

out from the contributions of pragmatic inference, both parties appeal to the notion of Fodor-modularity, and in particular to the notion of encapsulation. Hence, both minimalists and Relevance Theorists find common ground in their effort to locate the semantics-pragmatics distinction in the modularity of semantic processes and the isotropy of pragmatic inference.

1.3.2.2.1 BORG'S MINIMALIST ACCOUNT OF SEMANTIC PROCESSING

According to Borg, "semantic understanding [is] an activity deserving a modular treatment alongside phonetic and syntactic interpretation" (2004: 85). As stated here, Borg's thesis concerns an activity. As noted above, there are a few notions of modularity. There is Fodor's (1983) performance-oriented notion of modularity, on which modularity is a feature of certain processes or processing systems. There is also an alternative, competence-oriented notion of modularity, one that is articulated by Chomsky (1971, 1986), on which a module is a body of information (perhaps innately specified) that underlies some capacity. As noted, the operations that underlie the actual use of a capacity may or may not be restricted to the influence of the proprietary information in a Chomsky-module. Additionally, we can also think of a module as a hybrid of a Fodor- and a Chomsky-module, such that a module is a restricted body of information plus some processes that operate over only that information. After drawing this three-way distinction, Borg makes clear that her claim involves the third, hybridized notion of modularity (2004: 74-80). Hence, part of Borg's thesis in her 2004 concerns the "processes for realizing a … cognitive function," a function that is

both semantic and narrowly linguistic, in senses of 'semantic' and 'linguistic' that I discuss here in more detail.

In this regard, Borg points out an apparent point of divergence between her view and Fodor's. Fodor, again, claims that semantics is not a level of linguistic description (1989: 419). However, according to Borg, "it doesn't seem that we are forced to abandon the idea that semantic interpretation is a proper level of linguistic description" (2004: 85). Borg suggests that a semantic theory might pair natural language expressions with meanings, "namely, the expressions in the language of thought (perhaps concepts) to which they attach" (2004: 84). On Borg's view, such a theory would be a description of a semantic component of the language faculty. The language faculty, for Borg, is a hybrid of Chomsky- and Fodor-modularity, "containing specialized bodies of knowledge and operations on that knowledge, dealing with phonetics, orthographics, syntax, and semantics" (o.c.). Hence, in saying that semantics might be viewed as a level of linguistic description, Borg's claim is only that a hybrid Fodor- and Chomsky-modular language faculty might include a component that pairs linguistic objects (words and phrases) with expressions of the language of thought (both as a piece of knowledge, and as a computational process that transforms certain inputs to certain outputs).

Hence, the disagreement here seems to be mostly terminological, as Borg acknowledges. For Borg, we can reasonably label the pairing of words with concepts as 'semantic,' even if words derive whatever representational content they might have from the concepts to which they attach. As noted already, this is

Fodor's point with the quote about semantics not being linguistic, properly speaking: concepts, not words, have their representational contents by way of somehow being connected (by way of asymmetric dependency, on his view) with things out there in the mind-external world (Borg 2004: 84-85). In the sense in which Fodor uses 'semantic' in denying that semantics is subject to linguistic description (Fodor 1989: 419), to be a semantic fact is necessarily to be a fact about a relation, one relatum of which is "out there." If words have representational content, this is so only derivatively; a proper theory of content ought to be pitched at the level of concepts and thoughts. This is the sense in which Fodor denies that word-to-concept pairings are 'semantic.'

In effect, then, the question is about how we wish to use 'semantic.' We can restrict its use to refer to relations between representations and the things they represent (for example, relations like reference, satisfaction, truth), and thereby restrict use of 'semantic theory' to referring to an account of those representational relations (Lewis 1970). However, so understood, a semantic theory is beyond the purview of linguistic description. Alternatively, as Borg suggests, we can use 'semantic' in the translational sense, such that a theory is semantic just in case it pairs a word with another representation, such as a concept. Hence, one can agree with Fodor that concepts stand in the truth-conditional (or "real") semantic relation to things in the world, and also agree with Borg's hypothesis that the human faculty of language consists of a translational semantic component that maps expressions of a natural language to expressions in the language of thought.

Note, again, that the language faculty, for Borg, is both Chomsky- and Fodor-modular. Hence, her claim that semantics is modular consists of both a competence- and a processing-based version of the hypothesis. The information stored in the language faculty, and in its modular subcomponents (to include the semantic component), is encapsulated in the sense that it is a finite body of information (perhaps innately specified) that is proprietary to the specialized capacity that it underlies. And the processes of the language faculty are also encapsulated in the sense that their operations are influenced only by the bottom-up input and by the information in the proprietary databases of the language faculty. My target in what follows is primarily the processing-oriented hypothesis, as already indicated; however, I return to this distinction between the two notions of modularity, and their relation to semantics and pragmatics, in sections 1.4.3, 2.1.1, and chapter 5).

Borg ultimately defends a view in her 2004 on which the semantic system is involved in "grasp of literal linguistic meaning," which in turn is understood in terms of grasping highly general or "minimal" truth conditions for sentences. And the system is modular, in both the Fodorian and Chomskyan ways of understanding modularity. Hence, the operations of the semantic component of the language faculty are encapsulated from top-down effects of general cognition:

> We grasp literal sentence meaning in a vacuum as it were, free from the vast range of other things we know. To grasp the literal meaning of a sentence it no more matters whether it is the more predictable one possible in some context or whether it is the most surprising; its meaning, and a competent addressee's ability to grasp this meaning, remains unchanged. (Borg 2004: 105-106; cf. 85; 260 ff.)

As part of the process of grasping literal sentence meaning, the semantic processing system also performs the function of translating monomorphemic words into the primitive expressions of a language of thought, or concepts (Borg 2004: 84). So, given as input 'cat,' the semantic processor retrieves CAT; given as input 'orange cat' the processor retrieves both ORANGE and CAT and combines them in some way. Given a sentence as input, these processes of retrieval and composition eventuate in a truth-evaluable thought—albeit a highly general one. Crucially, as indicated in the above quote, these semantic processes are encapsulated from the top-down influences of general cognition. Borg's thesis, then, is an instance of the encapsulation hypothesis for semantics, one according to which the process of retrieving a concept given recognition of monomorphemic linguistic input is not influenced by extra-linguistic information.

1.3.2.2.2 Relevance Theorists

A very similar version of the encapsulation hypothesis for semantics is also defended by some Relevance Theorists. The core commitment of Relevance Theory is that communication in general is geared toward optimal relevance (Sperber & Wilson 1986). In what follows, I set aside most of the apparatus of Relevance Theory, focusing narrowly on the way in which the approach understands the distinction between semantics and pragmatics. As part of the process of utterance interpretation, Relevance Theorists posit two distinct components, one semantic and the other pragmatic. The semantic component is responsible for 'decoding' linguistic input. The pragmatic component is

responsible for drawing inferences about speaker meaning, in part on the basis of the deliverance of the decoding system, but also in large part on the basis of information available from non-linguistic sources, to include information about the immediate physical scenario, background knowledge, etc. (Sperber & Wilson 1986).

In characterizing the distinction between the semantic decoding system and the pragmatic system, Relevance Theorists commonly appeal to the more general distinction between modular and central processes, as the latter distinction is drawn by Fodor.

> The distinction between linguistic semantics and pragmatics is seen in performance terms, as closely tied to a distinction between two types of processes involved in understanding utterances: linguistic decoding and pragmatic inference … The decoding process is performed by an autonomous linguistic system, the parser or language perception module. Having identified a particular acoustic (or visual) stimulus as linguistic, this system executes a series of deterministic grammatical computations, or mappings, resulting in an output representation, which is the semantic representation, or logical form, of the sentence or phrase employed in the utterance. *It is a structured string of concepts*[.] (Carston 2002b: 57; emphasis added)

As antecedents for this conception of the semantics-pragmatics distinction, Carston cites Sperber and Wilson (1986), who explicitly note that the sense in which they take semantic decoding to be modular is Fodor's:

> The linguistic decoding system has all the hallmarks of automatic, reflexive perceptual systems such as hearing and vision. In the terms of Fodor (1983) … it is an input system rather than a central processing system [.] (Sperber & Wilson 1986: 177; cited in Carston 2002b: 57 and in Borg 2004: fn. 53)

So, at least for some Relevance Theorists, semantic processing is the work of a modular system in the sense of Fodor 1983. The inputs to this semantic decoding system are linguistic stimuli (words and phrases that are identified by other perceptual systems for phonological and syntactic processing); the outputs are concepts (given monomorphemic input) or structured strings of concepts (given complex input). The function of the semantic decoding system is to map words to concepts and to combine those concepts. And these two functions—concept retrieval and composition—are (by hypothesis) encapsulated in Fodor's sense.

Hence, when we focus on the processes that underlie utterance interpretation, the semantic minimalist and Relevance Theorist have a lot in common. Both agree that semantic processes are formally driven processes over narrowly linguistic input. Given a monomorphemic word as input, the semantic system delivers as output a concept. Given a phrase as input, the semantic system retrieves the concepts encoded by the constituent words and composes them, and in such a manner that neither retrieval nor composition varies as a function of extra-linguistic context.

A main point of contention between the minimalist and the Relevance Theorist is whether, given a sentence as input, semantic processing eventuates in something truth-evaluable. For the Relevance Theorist, semantic processing given a sentence as input rarely if ever results in construction of a truth-evaluable thought. Instead, semantic processing of a sentence often results in a propositional schema, one that can be enriched via pragmatic processes on the way to speaker meaning attribution. Minimalists, by contrast, maintain that semantic processing

of a complete sentence, in absence of (for example) attentional distraction, always eventuates in a truth-evaluable proposition, albeit a highly general one.

Consider (7), for example (due to Bach 1994[10]).

(7) Steel isn't strong enough.

Much like sentence (2) above, it seems that (7) fails to uniquely determine a proposition. (Strong enough for what?) In light of such examples, Relevance Theorists accept that semantic decoding of a sentence does not always eventuate in a truth-evaluable proposition or thought, but instead results in recovery of a propositional schema (e.g., Sperber & Wilson 1986; Carston 2002b). For the minimalist, semantic processing given a sentence as input always eventuates in a truth-evaluable thought, or a mental representation of the truth conditions for the sentence in question. Such thoughts (or truth conditions) are minimal in the sense that they are highly general and infrequently (if ever) meant by speakers. Hence, the minimalist meaning of (7) is the highly general proposition that there is something for which steel is not strong enough. According to the encapsulation hypothesis for semantics, recovery of such minimal contents is effected by an encapsulated processing system, not by way of abductive inference about speaker meaning.

Note, however, that from a processing-oriented perspective, such differences between minimalism and Relevance Theory concern the outcome of narrowly semantic processing of a sentence. When it comes to the processing of a

[10] It should be noted that Bach is not a Relevance Theorist. But he does argue for one conclusion that Relevance Theorists also endorse: that there are complete, indexical-free sentences like (7) that fail to determine a proposition.

monomorphemic word like ‘cat,’ there is apparently little difference between the two views. Both Borg and the Relevance Theorists appeal to a Fodor-modular, encapsulated semantic processor that translates monomorphemic words to atomic concepts. Hence, both minimalists and Relevance Theorists (at least as described here[11]) agree that, whenever a monomorphemic word has been recognized, its language of thought translation (an atomic concept) is retrieved in a way that is mandatory, rapid, and insusceptible to modulation by extra-linguistic context.

1.4 COGNITIVISM

The proposed connection between semantic theory and modular processing rests on several assumptions, some of which I have made clear in the above, some of which I have not. Those who do not share these assumptions may find the proposed connection between semantic theory and the notion of Fodor-modularity to be wrongheaded. In the present section, I articulate a framework for linguistic theory that I call ‘cognitivism.’ As the label is meant to suggest, the view is closely related to what is called ‘cognitive pragmatics’ (Sperber & Wilson 1986; see Carston 2002b for an overview). I characterize cognitivism in terms of a set of commitments or theses, to include at least a few of the following. The cognitivist perspective consists of a mentalist ontology for language; an internalist

11 Some Relevance Theorists propose a translational semantics that maps natural language expressions not to concepts or thoughts, but to constraints on expressible/retrievable concepts or thoughts (e.g., Carston 2012). The motivation for this adjustment of the theory is in large part consideration of polysemy. Given such an adjustment of the Relevance Theoretic approach, the Relevance Theorist and minimalist part ways concerning processing of monomorphemic inputs. For discussions of this idea outside of the Relevance Theoretic framework, see also Harris 2020 and Pietroski 2018 (though the latter is pitched at the level of competence, not performance). See overview in section 1.5 below and more detailed discussion throughout what follows, especially in chapters 4 and 5.

conception of semantics; an evidentiary understanding of the semantics-pragmatics distinction; and a focus on a description of language use (typically from the perspective of the hearer, rather than the speaker). This framework invites an approach to semantics and pragmatics that seeks to ground the distinction in facts about processing systems and their interactions during utterance interpretation.

1.4.1 MENTALIST LINGUISTICS; INTERNALIST SEMANTICS

Cognitivism involves a commitment to a mentalist ontology for language and an internalist, translational semantics. For the mentalist, linguistics is ultimately a branch of psychology (e.g., Chomsky 1965). To study language is to study a capacity of human minds, the way in which that capacity is acquired, and the way in which it is deployed during language use (both in the production and in the comprehension of linguistic utterances). For the internalist, part of this underlying linguistic capacity is semantic, in that it serves to link up natural language expressions with other mental representations (e.g., Chomsky 1964). From this mentalist and internalist perspective, to study linguistic meaning is to study systematic translation from natural language to another representational system in the mind.

Of course, other linguistic ontologies are on the market. Some prefer a Platonist or "realist" ontology for linguistics, on which linguistic entities are abstract objects and languages are complex systems, akin to mathematics, consisting of abstracta and various relations among them (see, e.g., Katz & Postal

1991).[12] Platonists are not primarily interested in questions about how individual human minds manage to process linguistic utterances, about whether certain components of that process might be encapsulated, or about whether the relevant notion of encapsulation is Fodor's. Such non-mentalist linguistic theories may of course appeal to different theoretical components of a linguistic theory (syntax, semantics, etc.), but their drawing such theoretical distinctions entails no commitments concerning the architecture of human minds. Answers to psychological questions about linguistic processing have very little directly to do with linguistic theory *per se*, as the latter is understood by the Platonist.[13]

For the mentalist, by contrast, linguistic questions ultimately reduce to psychological questions, and linguistic facts to psychological facts. So, the connection between linguistic meaning and modularity is likely to be of more interest to mentalist linguists than it is to Platonists. That said, mentalists often take the primary subject matter of linguistics to be an underlying capacity, or linguistic competence, idealizing away from various facts about the way in which this competence is deployed during language use, to include facts about the moment-by-moment processing of linguistic input. According to the typical prioritization for mentalists, linguistic use or performance is studied only

12 For an overview of different approaches to linguistic ontology, see Rey 2020: chapter 2.
13 However, the non-mentalist should not therefore conclude that the present dissertation is of no use to him. Louise Antony (2003) describes a case in which archeologists found clay pots in New Mexico that, along with an independent report by a Chinese astronomer, supported the conclusion that an astronomical event was observable from earth in AD 1053. (Cf. Carruthers 2003.) It seems less farfetched to expect that we might find psycholinguistic evidence to support claims about linguistic meaning, even if we construe languages Platonistically or otherwise abstractly. This point is argued for at greater length in Stainton 2009. What I wish to emphasize here is that (as Antony argues) psychological data might turn out to be relevant to the linguistic theory, even if one thinks that linguistic theory is not primarily about psychology.)

secondarily, as a source of evidence regarding the underlying capacity, which (again) is the primary subject matter. So described, mentalism is given one of its clearest characterizations in chapter 1 of Chomsky's (1965) *Aspects of the theory of syntax*. I return to this distinction (between competence and performance) and its relation to the modularity question in section 1.4.3. In any case, the cognitivist framework within which I approach questions about meaning and modularity is one that assumes a mentalist ontology for linguistics.

In addition, cognitivism includes a commitment to semantic internalism. An internalist semantics is, in effect, a translational semantics within a mentalist linguistic theory. For the internalist, meanings are mental entities. As Pietroski puts it, in a condensed paraphrase of a view articulated by Chomsky (1964, 1965, 1977, 1995, 2000), internalist meanings are "generated by and internal to human minds" (Pietroski 2017: 1). One simple version of an internalist semantic theory for natural language is the picture described above in connection with Borg and the Relevance Theorists. On this version, the meanings of words are concepts, and the meanings of phrases are combinations of concepts.

There is a well-known argument in philosophy of language against an internalist conception of meaning, due to Hilary Putnam (1975). Putnam assumed that, whatever the meaning of a word is, it (the word's meaning) must determine an extension, or the set of things to which the word applies. However, as Putnam argues, whatever determines the extension of a word, it (the extension determiner) cannot be "in the head," or psychological. From these two claims (that meanings

determine extensions, and that extensions are determined non-psychologically), it follows that meanings are not psychological entities.

Putnam simply assumes the premise that meanings determine extensions. The claim that extensions are not determined psychologically (that is, that membership in the set of things that 'water' correctly describes is not decided by psychological facts) is based on Putnam's "twin earth" thought experiment. Putnam asks us to imagine a planet that is just like earth, but where all the H_2O is replaced with some other molecule, XYZ. The judgment that Putnam reports is that, in such a scenario, our (regular earth word) 'water' does not correctly describe the watery stuff on twin earth, XYZ.

However, since (by stipulation) my twin earth counterpart and I are alike in every respect, including the facts about our minds (the only difference between our worlds being the molecular make-up of the watery stuff we drink, etc.), it follows that nothing psychological is responsible for determining the extension of 'water.' On Putnam's view, the extension of 'water' is instead determined by non-psychological facts (Putnam 1975).

As Chomsky notes, Putnam's argument here is invalid.[14] It is true that word meanings, whatever they turn out to be, *cannot both* be psychological *and* the things that determine extensions. However, it does not follow that meanings are non-psychological. After all, what follows from the negation of a conjunction is not a negation, but a disjunction of negations. If meanings are not both

14 For an overview of Putnam's argument and Chomsky's commentary, see Pietroski 2017, to which the present exposition is indebted.

psychological and extension-determiners, then either meanings are not psychological, or meanings do not determine extensions.

Hence, for an internalist approach to semantics, on which meanings are fundamentally psychological entities, there is no expectation that the meanings paired with words will determine the sets of things to which those words correctly apply. Indeed, internalists like Chomsky often point to considerations that suggest that efforts to specify extensions for words are misguided. (In this connection, recall the discussion of 'book' and '$bank_1$' above. What is in the extension of 'book,' the pulpy folios or the stories inscribed therein? See chapter 4.) In what follows, I largely take for granted that, as part of a mentalist linguistics, we can pursue an internalist semantic theory that pairs words with psychological entities of some kind, setting aside the question as to how such internalist word meanings might relate to objects in the mind-external world.

1.4.2 Evidentiary understanding of semantics-pragmatics

For the cognitivist, the semantics-pragmatics distinction is a distinction about evidence.[15] The goal of utterance interpretation is to infer the communicative intentions of a speaker. Speakers expect hearers to infer their intentions in part by recognizing the expressions they choose to utter, but also based in large part on knowledge of various non-linguistic facts. That is, I expect my audience will rely

15 Regarding this evidentiary conception of semantics, on which the linguistic meaning of uttered natural language expressions provides evidence for attributions of speaker meaning, see (for example) Sperber and Wilson 1986; Carston 2002b; Schiffer 2003; Neale 2004, 2005; Harris et al. 2017; Harris 2020. (Note, however, that the overall cognitivist perspective I describe here is not one that all of these theorists would accept as a whole.)

on the fact that I uttered the meaningful expressions that I uttered as a piece of evidence regarding my communicative intentions. From your perspective, the linguistic input provides some evidence about my communicative intentions. But I do not expect my audience to rely on such linguistic evidence alone. I also expect my audience to rely on evidence made available via extra-linguistic means, including background knowledge, awareness of the immediate physical scenario, etc. As discussed above, I expect my audience to rely on both sorts of information (linguistically provided and extra-linguistically available) not only in working out what I mean to indirectly communicate (recall Eck and 'Shoes!'), but also in deciding what I mean to communicate in a relatively direct fashion (recall 'book').

Hence, the cognitivist approach to the semantics-pragmatics distinction fits nicely with what is often called an 'intentionalist' approach to illocutionary speech act content. Intentionalism individuates illocutionary speech acts in terms of their intended effects. For example, assertion is an illocutionary speech act in which the speaker intends to get her audience to believe the proposition asserted and intends moreover for the audience to respond in this way in part on the basis of their recognition of the speaker's communicative intention (see, e.g., Grice 1957, 1968; Bach & Harnish 1979; Neale 2004). In many ways, my question in what follows concerns the hearer's perspective within this intentionalist picture. To infer the speaker's communicative intention, a hearer will rely on a mixture of evidence from linguistic and non-linguistic sources.

The encapsulation hypothesis for semantic processes in effect grounds, in a deep psychological division, the evidentiary version of the semantics-pragmatics distinction at the heart of intentionalist pragmatic theories. According to the encapsulation hypothesis for semantics, some of the evidence that pertains to inferences about speaker meaning is made available by a dedicated semantic processor that is impenetrable by higher-level cognition, and which maps natural language inputs to mental entities, perhaps expressions in a language of thought. In other words, the encapsulation hypothesis for semantics naturalizes the distinction between linguistic and extra-linguistic evidence, and likewise naturalizes questions about how the two types of information interact during utterance interpretation. A dedicated linguistic processing system includes a semantic component that serves to translate linguistic inputs to composable mental entities (for example, concepts), where the latter play an evidentiary role in the broader inferential (or "central") process of utterance interpretation.

1.4.3 Performance-oriented and hearer-oriented approach

Again, in what follows, I focus mostly on linguistic performance, and only indirectly on the competence that underlies this performance. Moreover, I focus on performance from the perspective of the hearer, not the speaker (that is, on comprehension, not production).[16] In focusing on processing, I follow the work of

[16] While the processing-oriented focus is constitutive of the cognitivist framework as I see it, disproportionate attention to the comprehension processes carried out by a hearer, as opposed to the processes of production on the part of the speaker, is not. Philosophers working in the cognitivist tradition have generally focused more so on the hearer largely because the hearer's task is a robustly *inferential* one, and the relevant inferences appear to be non-deductive or ampliative. Hence, disproportionate focus on the perspective of the hearer, at least among philosophers of

Relevance Theorists. As Carston notes, "the relevance-theoretic approach has probably the strongest cognitive *processing* orientation of all the various post-Gricean strands of pragmatic work" (2002: 12, fn. 2; original emphasis).

As mentioned, I do not commit to or even rely on certain core aspects of Relevance Theory in what follows. For example, I do not assume or defend here the view that utterance interpretation is driven entirely by a presumption of optimal relevance. However, my explanatory goals are ultimately the same as the Relevance Theorist's: to make some progress in explaining the processes and mechanisms that underlie utterance interpretation. That said, my goal, again, is to make only a modest contribution in this regard, by assessing efforts to locate the boundary between semantic and pragmatic contributions to utterance interpretation in the modularity of semantic processes. (As indicated already, and as discussed further in section 1.4.4 below, the goal is yet more modest, as I focus on semantic processes whose inputs are monomorphemic words).

The use-orientation of Relevance Theory is prefigured in the work of so-called ordinary language philosophers. In brief, here is the relevant history (guided by Carston 2002b). Formalist, "ideal language" philosophy—as pursued by Frege, Russell and ("early") Wittgenstein—was meant to offer an explicit characterization of the inference patterns that are valid, or truth-preserving. This project in turn requires that we characterize the logical forms of propositions. And the sentences of a natural language can serve as a guide in this regard, but only an unreliable one. Discerning the logical forms of propositions requires that we

---

language working in the tradition in question, is arguably an incidental side-effect of a more general epistemological orientation within philosophy.

abstract away from many considerations about natural languages. ‘Nobody knows’ and ‘Sam knows’ initially seem to have the same logical form, but that is only because natural languages are messy. Hence, in characterizing the structure of propositions, it is often preferable to deploy a “non-natural” or artificial language, such as a predicate calculus.

As a reaction to ideal language philosophy, so-called “ordinary language” philosophers—for example, Austin, Ryle and Strawson—took as their primary subject matter the meanings of the natural language expressions that we ordinarily use. Initially, ordinary language philosophy was considered an alternative to formal or ideal language philosophy. So, for example, it was sometimes suggested that, to characterize the meaning of a natural language expression *just is* to describe the uses to which it is put by speakers, there being nothing more to say than that (see, e.g., Wittgenstein 1953: section 43). From this viewpoint, any claim that purports to be a claim about what a word means on its own, independently of its uses, is spurious.

However, following the work of Paul Grice (e.g., 1967), the two approaches are now generally seen as complementary. A formal semantic theory specifies the meanings of natural language expressions themselves, while a pragmatic theory gives an account of the uses to which speakers and hearers put such expressions. Grice was motivated to preserve parallels between logical and natural languages that had been worked out by Frege, Russell and others—for example, to preserve the hypothesis that the natural language expression ‘and’ shares a meaning with the truth-functional connective ‘&.’ We can use ‘and’ to communicate a claim

about temporal sequencing (‘got into bed and took of his shoes’), or causation (‘fell off the table and shattered’), or plain-old truth functional conjunction (‘tired and hungry’). However, this should not lead us to conclude that ‘and’ is in fact three words, or that ‘and’ itself has three different meanings, or that to give the meaning of ‘and’ *just is* to list these uses. As Grice emphasized, speakers routinely exploit, to communicative effect, the fact that a hearer will try to rationalize their actions (Grice 1967). Hence, what gets communicated always involves a massive contribution of general rationality, and not just an appreciation of the sentence uttered and its meaning. Hearers rely on the information provided by the sentence uttered itself, but also rely on background knowledge, awareness of the physical scenario, considerations about the intentions and goals of interlocutors, and so on. In light of this observation, and by appeal to certain general principles that govern rational communication, Grice offered an explanation for the various uses and understandings of ‘and,’ an explanation that preserved the idea that ‘and’ shares a truth-functional conjunctive meaning with ‘&.’ In short, while ‘and’ and ‘&’ share a meaning, speakers manage to indirectly communicate temporal and causal messages by relying on the inferential capabilities of hearers (Grice 1975).

In large part, Gricean pragmatics as originally developed was meant to reduce homophony, or to solve some other problem in natural language semantics (whether ‘it looks red’ entails that it is not red, whether ‘the dog barked’ entails that there is only one dog, etc.). The pragmatic processes underlying recovery of communicative intentions are not Grice’s primary explanatory target. Pragmatic

considerations are generally appealed to only to establish what is and what is not in the purview of semantic theory, the latter being the primary focus. On this perspective, pragmatics has been caricatured as a "wastebasket" for semantics (Bar-Hillel 1971: 401), a discipline whose primary task is to collect facts that are unfit for semantic explanation.

'Cognitive pragmatics' shifts the focus from pragmatics as an adjunct to semantic theory to pragmatics as a proper object of study in its own right (Sperber & Wilson 1986). On this approach, pragmatics is a capacity of the mind, a system for interpreting communicative behavior, to include linguistic utterances. The goal for cognitive pragmatics is not rational reconstruction of how a system or mind might go about inferring communicative intentions, but an understanding of the actual processes and mechanisms that underlie moment-by-moment utterance interpretation.

On this point, compare Jennifer Saul's (2002) notion of an 'audience implicature' versus a genuinely Gricean implicature. The latter is a normative notion that involves abstraction away from actual processing as it occurs in hearers' minds. Whether a speaker has implicated that Q by saying that P in the normative sense depends in part on whether it is *rational* for the speaker to expect the hearer to recognize the speaker's intention to communicate Q by uttering P. By contrast, whether Q has been "audience implicated" in the non-normative sense depends only on whether the hearer *in fact* attributed to the speaker the intention to communicate Q. Cognitive pragmatics aims to give an account of how such non-normative audience implicatures are computed by hearers.

As part of most accounts in cognitive pragmatics, some distinction is drawn between semantic decoding and pragmatic inference, as discussed above in connection with Relevance Theory and the issue of modularity. In effect, my project in what follows is to ask, from the perspective of cognitive pragmatics, whether the decoding function is carried out by a Fodor-modular processing system, and what that would entail regarding linguistic meaning. It is in this sense that I have suggested that we might "naturalize" the semantics-pragmatics distinction (cf. Quine 1969).[17] The cognitivist's goal, in effect, is to understand how Saul's audience implicatures are actually calculated by hearers during the moment-by-moment processing of linguistic input. My more modest goal in what follows is to ask how the notion of semantic modularity might contribute to such a theoretical understanding.

So, as I use the term, the cognitivist is a mentalist linguist whose goal is an explanation of the processes and mechanisms involved in utterance interpretation. However, for mentalists (as noted above), the primary subject matter of linguistics is generally not considered to be the psychology of performance, or the ways speakers and hearers use language, but instead the knowledge or competence that underlies that use (Chomsky 1965). Again, my attention here is restricted primarily to questions about performance or language use, and in particular the processes that underlie a hearer's interpretation of a speaker's utterance. So, the

[17] It should be said there that I in no sense intend to argue that such a "naturalized," descriptive approach to utterance interpretation should *replace* the normative inquiry. Nor do I mean to suggest that normativity is incompatible with a general commitment to naturalism. I use the term 'naturalized' here only to indicate that I pursue a descriptive question about semantics and pragmatics, instead of a closely related normative question.

operative notion of a module is Fodor's, not Chomsky's. That is, my focus here is the hypothesis that there is a cognitive subsystem that is dedicated to the moment-by-moment processing of linguistic input, a subsystem that can in some sense be called 'semantic,' in that it mediates between identification of linguistic input and attribution of speaker meaning, mapping a word or phrase onto a representation (for example, a concept) that plays some further role in utterance interpretation. That said, I do not focus on performance for its own sake; ultimately, my hope is that the study of language use will afford insight into the structure of linguistic knowledge.

But any such insight will have to be achieved cautiously. Of course, a theory of competence does not by itself entail a theory of performance. An account of semantic competence does not carry with it any specific commitments about how that competence is deployed. So, we cannot expect to find any direct reflection of a theory of competence in what we observe of performance. As Chomsky notes, we could only expect performance to directly reflect competence by idealizing away from various factors that inevitably obtain in any real speech situation, such as a speaker or hearer's "memory limitations, distractions, shifts of attention and interest, and errors" (Chomsky 1965: 3), as well as "false starts, changes of plan mid-course, and so on" (o.c.: 4). It follows, then, that one might be committed to Chomskyan modularity in a theory of semantic competence but reject (or be agnostic about) the hypothesis that there is a Fodor-like input-system that deploys this competence during language use. Indeed, a competence theorist might ask

what reason we have to expect to learn anything about competence via the study of performance.

That said, we cannot ever directly observe competence, semantic or otherwise, as Chomsky notes (regarding grammatical competence and performance):

> The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance. Hence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior. Mentalistic linguistics is simply theoretical linguistics that uses performance data (along with other data, for example, the data provided by introspection) for the determination of competence, the latter being taken as the primary object of its investigation. (o.c.)

It follows that, if we hope to empirically test a competence theory, we need to find some (perhaps highly) indirect reflection of the theory's categories and distinctions in what we observe in performance. Without such an indirect reflection, we have no hope of finding data to guide our competence theory. So, while a competence theorist is right to ask, skeptically, exactly how we should expect to learn about competence by studying performance, it is unclear how else we could make progress in this regard. Moreover, as Chomsky acknowledges, while his theory of grammatical competence does not entail any commitments about processing mechanisms, it does help us "with determining abstractly the properties that such mechanisms must exhibit and the functions they must perform" (o.c.).

By ‘performance’ I do not mean to exclude considered judgments of competent speakers about (e.g.) acceptability, felicity, truth, entailment, etc. Such judgments are themselves the product of performance systems,[18] subject to the vicissitudes listed in Chomsky 1965, and so not a direct reflection of linguistic competence. However, I largely focus, in what follows, on performance measures that are less intrusive than judgments of truth or felicity. In general, I look to controlled experimental studies using behavioral or neural measures, some of which are taken to reflect felicity judgments, albeit in a very indirect way that requires the mediation of linking hypotheses. For example, comparison of the relative degrees of electrical activity recorded on the scalp in time-locked connection to linguistic stimuli may afford one promising route for probing hearers’ judgments of the felicity of an unfolding sentence (Kutas & Hilyard 1980), but only if we articulate plausible linking hypotheses that connect the electrophysiological observations with different aspects of sentence comprehension (see chapter 3).

So, the present inquiry should be of interest to proponents of both Chomskyan and Fodorian modularity for semantics, although the relevance of this discussion for the Chomskyan will be considerably less direct. Exactly how observations about processing relate to a competence theory depends on the

18 See, e.g., Drożdżowicz on the ‘voice of performance’ (2015: 105, cited and discussed in Rey 2020: chapter 8). While the notion is applied regarding syntactic competence and performance, the same point applies regarding semantics: semantic competence is only ever indirectly reflected through performance systems. Reflective judgments about grammaticality, felicity, truth—all of these are the result of the operation of performance systems, contrary to the suggestion implicit in the Chomsky quote above.

linking hypotheses we adopt—hypotheses that state how observed differences in the processing profile relate to posited distinctions in the competence theory.

1.4.4 NARROW FOCUS ON MONOMORPHEMIC WORDS

Much of semantic theory is focused on issues of meaning composition, or construction of the meanings of phrases and sentences from the meanings of their constituents. I focus instead on what you might think of as the "smallest" semantic operation: retrieval of the most basic (or primitive) units for semantic composition. Compositional theories generally allow for composition to "bottom out" at the level of its most basic units. And, so, part of the semanticist's job is to say what those primitives are.

The sorts of things I loosely called 'contents' above in section 1.1 are typical candidates for the semantic primitives (referents, concepts, truth-conditional contributions), though there are others, to include mental entities that place constraints on retrievable contents (as discussed throughout what follows, especially in chapter 4). In any case, I mostly set aside consideration of semantic composition and focus narrowly on processes that appear to be non-compositional, but potentially semantic nonetheless, in that they may involve recovery of such basic units for composition.

My focus on monomorphemic words is not constitutive of the cognitivist framework. However, while it is generally agreed that mode of semantic *composition* is context-invariant, it has been widely suggested that the concept or meaning that you *retrieve* given a word as input is contextually variant (e.g.,

Searle 1980, Travis 1985, Carston 2002a, Recanati 2004). Hence, when it comes to semantics, the substantive and more controversial modularity question concerns word-level processes.

1.4.5 EXPERIMENTALISM

Given its processing-orientation, a cognitivist account of the distinction between semantics and pragmatics is potentially answerable to evidence from a wide range of sources. The potentially relevant sources of evidence include those that semanticists have typically relied on—reflective judgments about truth, entailment, and felicity. But the cognitivist is also answerable to evidence from a wide variety of other sources, to include experimental investigation of language acquisition, sentence comprehension and sentence production, and studies of linguistic impairment (cf. Carston 2002b: 4). In what follows, I discuss the encapsulation hypothesis for semantics in light of a wide range of experimental methods and results in the study of adult sentence comprehension.

More generally, the cognitivist perspective I adopt throughout what follows is committed to a robust experimentalism. That is, the perspective commits one to expecting the methods and results of psycholinguistic experimentation to be potentially relevant to semantic and pragmatic hypotheses. This experimentalist commitment is of course both optimistic and cautiously so. After all, there is no guarantee that the currently available experimental psycholinguistic methods and results will in fact shed any light on the relevant issues, to include even the narrowly semantic contribution of 'book.' However, the cognitivist viewpoint

commits one to expecting such efforts to be at least in principle potentially relevant—and perhaps especially so, among all the possible sources of evidence—to questions about the semantic and pragmatic processes underlying utterance interpretation.

1.5 What does a semantic processor do?

From the perspective just characterized, I ask in what follows whether semantic processes are encapsulated, and if so in what sense. It remains to be said what processes might be semantic in the relevant sense. One such function is assignment of an index to a word-form (for example, going from /bæŋk/ to '$bank_1$'). However, the function that I focus on primarily involves assignment of a content to an indexed word-form (for example, going from '$bank_1$' to a concept). I start here by describing index assignment in more detail, then turn to content resolution. For both functions, the goal in this section is only to characterize the functions and briefly motivate why you might ask (i) whether they are carried out by Fodor-modular, encapsulated processing systems, and (ii) what this would entail for semantic theory, conceived mentalistically.

Index assignment is paradigmatically involved in homophony resolution. Given recognition of an instance of the phonological word-form /bæŋk/, the function of index assignment is to assign a feature, either feature [1] or feature [2]. Assignment of [1] yields '$bank_1$.' Assignment of [2] yields '$bank_2$.' What differentiates '$bank_1$' and '$bank_2$' is not their pronunciation, since they share the same pronunciation. Neither are '$bank_1$' and '$bank_2$' distinguished syntactically,

since the two words share the same syntactic distribution. What, then, differentiates 'bank$_1$' and 'bank$_2$'? What difference do the diacritics register? From a processing-oriented perspective, it is hard to answer this question without appealing to concepts (or some other content-like or content-adjacent entity). To assign [1] it is to constrain in one way the possible concepts with which its host might be paired (where at least FINANCIAL INSTITUTION and BUILDING satisfy the constraint imposed by assignment of [1]). To assign [2] is to constrain in another way the possible concepts with which its host might be paired (where RIVERBANK satisfies the constraint imposed by assignment of [2]). In this sense, [1] and [2] might be considered semantic features: they serve to constrain possible pairings with contents, or possible meanings.

Why think that index assignment might be encapsulated? After all, one might instead suppose that what is needed for index assignment is a reliable enough degree of accuracy in a task for which the space of options is very limited. What matters is that, given recognition of /bæŋk/, the linguistic processing system does not end up assigning [3] or [88], but assigns either [1] or [2], and with a fair degree of accuracy. Consider the following example. Suppose that a hearer encounters the partial sentence (8).

(8) I needed to cash a check, so I went to the _____.

And suppose that, at the point of continuation following 'the' and marked by the blank space, the speaker begins to pronounce /bæŋk/. Does the hearer's mind even consider assigning [2] (which, again, is the feature that constrains possible content pairings to the fluvial, not financial) at any stage of (8)'s processing? If so—if

there is a stage at which the hearer's mind considers assigning both [2] and [1] as the index for the unfolding continuation—then one might take this to show that index assignment is encapsulated in the Fodorian sense. If, despite the fact that the preceding context strongly biases assignment of [1], a hearer's mind considers assigning [2] to the unfolding continuation of sentence (8), this would suggest that the index assigning system is (at least at certain stages) encapsulated from top-down, extra-modular influences.

A lot of experimental work has been done along these lines. As it turns out, most stages involved in word recognition appear to be susceptible to some degree of top-down influence, leaving very little room for the kind of strictly encapsulated processing that Fodor characterizes. In a meta-analysis of priming experiments meant to study whether word recognition was influenced by extra-modular factors, Margery Lucas reports that "about 3/4 of all observations and 4/5 of all studies showed an appropriateness effect of some kind," (1999: 387) i.e., a top-down effect of extra-linguistic context on the process of recognizing a word-form. As Lucas notes, even a small effect size is sufficient to disconfirm the Fodor-style modularity hypothesis (Lucas 1999: 386).[19]

Perhaps, however, what matters is not whether a process is encapsulated in the strict, Fodor-style sense. Perhaps a sufficient degree of encapsulation of the limited sort proposed by so-called "massive modularists" would suffice to ensure tractability of the process of (say) word recognition. One of the recommendations

[19] See more recent evidence in support of this conclusion in Gaston 2020, but with an eye to the specific mechanisms underlying contextual influences on word recognition.

I make in what follows, concerning the semantic modularity hypothesis, is that such non-Fodorian notions of modularity might fit better with the available data.

These results are of course subject to debate, and there are proponents of Fodor-modular models of word recognition (e.g., Forster 1979). There may be a certain amount of encapsulation of some stages of word recognition. Suppose, for the sake of argument, that we were to conclude that there is an encapsulated stage of word recognition, a stage at which a system responsible for index assignment generates the possible indices for assignment, prior to a further process of selection of one of the candidates, and where the process of generation (if not selection) is insusceptible to top-down, extra-modular influences of pragmatic considerations. What would this entail for a semantic theory, conceived mentalistically?

One might take the encapsulation of index assignment (or, at least, a component process thereof) to ground the distinction between semantics and pragmatics. However, many cognitivists would be unsatisfied with such a conclusion. As shown above, most cognitivists endorse a picture of semantic theory on which it is in the business not just of recognizing the homophony of /bæŋk/ (and generating the optional resolutions) but is instead or additionally in the business of pairing the results of homophony resolution with an expression in the language of thought, or a concept. Hence, another function that might be both semantic and encapsulated (at least in part) is the process of concept retrieval itself, or translation by the semantic system of a monomorphemic, indexed word-form to an atomic concept in the language of thought. Since this is the kind of

potentially semantic process that philosophers of language have been primarily interested in, it remains my primary focus in what follows.

In the simplest case, content resolution would be a function that maps an indexed word-form to a concept—for example, mapping 'cat' to CAT. So understood, the function in question is not aptly referred to as 'resolution,' since the output would be uniquely determined by the input in a one-to-one fashion. And, so understood, there is little reason to ask whether such a process is encapsulated or not. Since there is no decision to be made, there is accordingly no decision that might be influenced by top-down, extra-modular pressures. However, polysemy shows that, in some cases, the indexed word-form stands in a one-to-many relation, not to a single content, but to several contents. For example, given an assignment of [1] and recognition of 'bank$_1$,' a decision between the institutional and physical senses of 'bank$_1$' has not yet been made. And, as I argued briefly above (and in more detail below in chapter 4), this decision is one that can be made *only* by appeal to pragmatic, extra-linguistic considerations.

What, then, could be the output of an encapsulated semantic processor, given as input an indexed word-form that is polysemous? This question is addressed in detail in chapter 4, where I contrast a view on which polysemes are mapped to chimerical combinations of concepts (e.g., Langacker 1991) to a view on which polysemes are mapped to "pointers" to vaguely defined regions of conceptual space (e.g., Carston 2002a). On both views, the immediate affordance of semantic processing given an indexed word-form is not a specific concept, but either a list of concepts or a pointer to a space of concepts. As I argue,

consideration of polysemy entails that an encapsulated semantic process over indexed word-forms cannot itself eventuate in retrieval of a single concept. Before considering such questions about what an encapsulated semantic processor might yield as output, however, I discuss in chapters 2 and 3 whether we should agree with Borg and the Relevance Theorists that semantic processes are Fodor-modular. In chapter 2, I briefly review a handful of arguments that have been offered in support of the claim that semantic processing (of the sort that goes beyond index assignment) is modular. Most of the considerations that have been offered in support of the Fodor-like hypothesis that semantic processes are encapsulated turn out to be unconvincing.

As usual, assessment of the hypothesis that semantic processes are encapsulated requires careful interpretation of various observations that appear potentially relevant to its confirmation or disconfirmation. Hence, in chapter 3, I look to a range of experimental methods and results in psycholinguistics that appear to be pertinent to assessment of the encapsulation hypothesis for semantics. This experimental record generally indicates that top-down influences of pragmatic considerations have rapid effects on even very early stages of utterance interpretation. As is the case for vision and other sensory processes, there appears to be a substantial empirical challenge for the proponent of the Fodor-like hypothesis that semantic operations are insusceptible to any top-down, extra-modular effects of general cognition. Background knowledge, hearers' beliefs about speakers' intentions—such pragmatic considerations have a very rapidly detected effect on utterance interpretation.

In chapter 4, I shift attention to the phenomenon of polysemy, which serves nicely as a case study for debates about modularity for at least two reasons. First, polysemy dissociates word recognition and concept retrieval, and so allows us to keep separate the two versions of the semantic modularity hypothesis (index assignment versus content resolution). Second, there is considerable debate about what a translational semantic theory might pair with a polysemous word (a privileged concept, a list of concepts, a pointer to conceptual space). Accordingly, in chapter 4, I ask which of these entities might be retrieved by an encapsulated word processor, and what evidence or considerations might adjudicate among those options.

As I ultimately argue, proponents of the idea that the encapsulation of semantics somehow grounds the semantics-pragmatics distinction are left with three options in light of the data and arguments I bring together here. They can articulate principled explanations of the relevant observations such that they are compatible with the strict, Fodor-like conception of encapsulation (as I demonstrate in chapter 3). Alternatively, proponents of semantic modularity might appeal to an alternative, non-Fodorian notion of modularity, one that involves only a limited form of encapsulation, less strict than Fodor's. Or one might of course be led to abandon the idea that encapsulation is the notion that serves to separate semantics from pragmatics in the communicating mind. These last two options, and their consequences regarding both a mentalistic semantics and the philosophy of language more generally, are discussed in the concluding chapter 5.

Chapter 2: Assessing the arguments for encapsulated semantics

Several arguments have been offered in support of the hypothesis that semantic processes are modular in a Fodor-like way, such that they are insusceptible to top-down, extra-modular influences of pragmatic considerations. In this chapter, I review those arguments in detail. I then show why they are, for the most part, unsuccessful arguments. I do so in rapport with parallel debates about encapsulation in other psychological domains, especially vision. Given the failure of most other strategies for its defense, the case in favor of encapsulation comes down to careful consideration of the experimental record. Since it is an empirical hypothesis, the encapsulation hypothesis for semantic processing must be tested against evidence, a project that I therefore pursue in the subsequent chapter. However, before doing so, it is necessary to convincingly set aside some commonly recurring arguments for the Fodor-modularity of semantics.

2.1 Overview

There are at least four types of argument offered on behalf of the idea that semantic processes are informationally encapsulated in the sense of Fodor 1983. The first argument is a kind of programmatic one, according to which the encapsulation hypothesis is the best (and perhaps the only) way to preserve the theoretical distinction between semantics and pragmatics within a cognitivist

account of utterance interpretation (see, e.g., Harris 2020). Another way in which the encapsulation hypothesis can been defended is by arguing that semantic operations exhibit other features of Fodorian modules; if this is so, we have good reason to conclude that semantic operations are encapsulated (such an argument appears to be implicit in both Borg 2004 and Harris 2020). A third way to argue for the encapsulation hypothesis is to point to psychological observations that appear to support it—e.g., by appeal to the semantic equivalent of visual illusions, such as the Müller-Lyer (Harris 2020: 7, borrowing from Pettit 2002: 544). Alternatively, the encapsulation hypothesis can be defended by pointing to the systematic absence of evidence against it (e.g., Harris 2020: 9).[20]

I'll call these, respectively, the 'programmatic argument,' the 'cluster argument,' the 'positive evidentiary argument,' and the 'negative evidentiary argument.' Having described and motivated each of these arguments below, I then go on to show why the first three are unsuccessful. In short, Fodorian encapsulation is not the only way to individuate mental operations, and therefore not the only way to secure the distinction between semantics and pragmatics in a psychological science of language use. So, the programmatic and cluster arguments are invalid; and, as inductive arguments, they are quite weak. Furthermore, familiar arguments concerning the persistence of illusory percepts are likewise invalid, since the observations relating to such illusions only show that extra-modular information cannot override entirely the bottom-up input, not

20 Throughout this chapter, Dan Harris' (2020) case for semantic modularity, along with that of Emma Borg (2004), serves as both a guide and a foil. There are many respects in which I am sympathetic to these arguments, but others in which I am less so, as I make clear throughout this chapter (and the next few).

that such information is incapable of influencing the modular processing in question (cf. Ogilvie and Carruthers (2016: 726) on vision).

Ultimately then, the case for (or against) the encapsulation hypothesis for semantics comes down to whether we have much in the way of evidence against it. This is unsurprising. With regard to perceptual and grammatical processing, Fodor himself went to considerable effort to defend encapsulation by raising objections to arguments for unencapsulation. Given some plausible guidelines for assessing putative instances of unencapsulation, once can explain away apparent counterexamples and preserve the encapsulation hypothesis for the domain in question (Fodor 1983: 72-85). What is therefore needed, as I argue, is a careful application of guidelines or ground rules for analysis of observations that might appear to be incompatible with the encapsulation of semantics. No serious defense of Fodor-modularity for semantic processing would be complete without discussion of apparent counterexamples (for which, see chapter 3). First, however, I show why the other arguments for unencapsulation are unsuccessful.

2.1.1 THE PROGRAMMATIC ARGUMENT

The programmatic argument rests on the idea that the encapsulation hypothesis is the best (and perhaps the only) way to preserve the theoretical distinction between semantics and pragmatics within a cognitivist account of utterance interpretation. Dan Harris, a proponent of the encapsulation hypothesis,[21] says this on its behalf:

21 Harris seems to be primarily focused on semantic composition, not the "smallest' semantic process of meaning retrieval. However, as discussed below, Harris' 'mud' thought experiment suggests that, on his view, the Fodor-modular semantic processor maps lexical input onto composable semantic units (Harris: 2020: 8; discussed in detail below). And Harris does say that

> One advantage of this view is that, unlike its competitors, it allows for a precise semantics-pragmatics interface that is grounded in underlying facts about human cognitive architecture. The semantics-pragmatics interface turns out to be an instance of the interface between central cognition and its peripheral input-output systems. If this is also where we locate the perception-cognition interface—an admittedly controversial view—then the semantics-pragmatics interface also turns out to coincide with the interface between language perception and general-purpose cognition. (Harris 2020: 13)

Elsewhere, Harris suggests the stronger claim that any adequate account of utterance interpretation must somehow preserve a sharp architectural distinction between semantic and pragmatic operations:

> On one hand, semantic competence informs the operations of the semantic module. On the other hand, our beliefs and other intentional mental states are drawn upon by pragmatic reasoning but are off limits to the semantic module. An empirically adequate account of the psychological states underlying language use must not conflate these two bodies of information. (Harris 2020: 6)

The best way to avoid conflation of semantic and pragmatic information, we are led to believe, is by way of a Fodor-modular—and, in particular, an encapsulated—semantic system.

Similar considerations are offered in support of encapsulation hypotheses in other cognitive domains, such as vision. In short, the idea has been that the distinction between perception and cognition carves a mental joint, and the

---

lexical knowledge is part of the database for the semantic module, on his view (Harris 2020: 6). There is room for one to take the position that composition is the only semantic operation, retrieval being pre-semantic. That is, one might propose that the database for the semantic module stores pairings of lexical items and their semantic "values" (constraints, on Harris' view, as discussed in detail below in chapter 3), but there is no Fodor module that context-invariantly maps lexical input onto its stored value. Strictly speaking, what Harris writes is compatible with this option. However, again, his 'mud' thought experiment indicates that, among the operations that he takes to be semantic, Harris includes some non-compositional processing of lexical, non-phrasal input.

distinction between encapsulation and unencapsulation affords "the most natural and robust distinction between types of mental processes" (Firestone & Scholl 2016: 1). It is this tradition that Harris appeals to in arguing that semantics must likewise be an encapsulated system if the distinction between semantics and pragmatics is to carve the mind at a natural joint (Harris 2020: 5). As briefly acknowledged above, this line of thought is attractive. We expect a psychological science of utterance interpretation to yield generalizations that carve psychological nature at its joints. And the encapsulation hypothesis for semantics therefore affords us a "natural and robust" way of grounding, within a science of language use, the distinction between semantic and pragmatic sources of information.

However, there are two main problems with the programmatic argument. First, it assumes that the distinction between semantic and pragmatic information should be preserved in an empirically adequate psychological science of utterance interpretation. But surely this is itself an empirical hypothesis—i.e., that the theoretical distinction between semantics and pragmatics is one that carves a natural joint in the communicating mind. Perhaps it does not, at least not in any way beyond the fact that there are these two types of information. We commonly assume that there is a component of utterance interpretation that is semantic in the sense that it maps words and phrases onto representations of one sort or another (perhaps truth-evaluable, perhaps not; perhaps highly specific, or perhaps instead highly general; see chapter 4 below). But this assumption may well turn out to be false. Perhaps there is, after all, no semantic waystation in the mind that mediates

between recognition of a word or phrase and the drawing of inferences about speaker meaning. Perhaps word recognition triggers activation of some associated concepts, none of which can be called the privileged meaning of the linguistic input. None of these options is ruled out a priori. So, the programmatic argument for the encapsulation of semantics is unsuccessful, unless we antecedently agree that semantics must be the study of an isolable component of utterance interpretation.

After all, one might maintain that semantic competence is an isolable component of the communicating mind (and so a module in the Chomskyan, competence-based sense), but be neutral, skeptical, or even hostile to the hypothesis that this competence is deployed by a special-purpose processing system that is isolated, somehow, from information and operations available or occurring elsewhere in the mind.

A crude analogy will serve to emphasize this point. We have good reason to maintain that black pepper and oregano are different things, and accordingly that to know the one is in some important respect different than what it is to know the other. However, we can imagine a performance system that ignores this distinction. Suppose there is a chef who grabs either pepper or oregano, tossing it in the pan or pot whenever a recipe calls for one or the other. That this chef, the performance system in our crude analogy, ignores the distinction between black pepper and oregano clearly does not entail that there is no such difference. Neither does the failure of this chef to distinguish between the two ingredients entail that he does not know they are different ingredients.

Somewhat less crudely, we can imagine a processing system that operates over both the sounds of birds and their physical shapes, yielding as output a judgment about species. Were we to somehow discover that this bird recognition system in effect ignores the distinction between these two sources of information, this discovery would not give us reason to reject the distinction between knowledge of bird sounds and knowledge of bird shapes (never mind the distinction between sounds and shapes).

Likewise, even if it were to turn out that the processes unfolding in the minds of language users effectively ignore the distinction between semantic and pragmatic sources of information, this would not falsify the semantic competence theorist's hypothesis that the hearer's mind includes a special database of semantic knowledge. And that special database would then provide the subject matter for a science of that competence. As mentioned above in clarifying the Fodorian focus of this discussion, the Chomskyan modularity hypothesis for semantics does not entail its Fodorian counterpart. And so, observations that make trouble for the Fodor-like modularity hypothesis for semantics do not *ipso facto* make trouble for the Chomskyan. Harris writes, "an empirically adequate account of the psychological states underlying language use must not conflate these two bodies of information," semantic and pragmatic (2020: 6). However, it could very well be that, for all we can tell from the available evidence, language *use* does conflate these two bodies of information. And if so, and if Harris is right that an empirically adequate account of utterance interpretation requires that we avoid

such a conflation, one might be tempted to conclude that an empirically adequate account of utterance interpretation is not in the offing.

In fact, such a line of thought is often explicitly endorsed (see discussion in Carston 2002b: 1-2, from which the citations here are drawn). Chomsky argues that a scientific theory of language use—a theory of how interpreters assign interpretations to speakers on an occasion of use—is not possible. As he argues, communication and interpretation are not topics that submit to scientific study, since the relevant theory would be a theory of everything (Chomsky 1992: 120; cf. Davidson 1986). As Fodor explains in more detail, focused (as I am) on the perspective of the hearer, interpretation is a matter of belief fixation (that is, fixing a belief about speaker intentions). Belief fixation is isotropic (any belief is potentially relevant to any other) and holistic (such that a belief is fixed only as a part of the set of one's beliefs taken as a whole). In Fodor's view, the isotropic and holistic inferential processes of central cognition are unlikely to be subject to fruitful scientific investigation (Fodor 1983, 1986, 1987, 2000). If Harris is right that an empirically adequate account of language use requires that we not conflate semantics and pragmatics, this insight—a conditional one, in effect—can be responded to either by affirming the antecedent and searching for ways to preserve the distinction (as Harris does) or by denying the consequent and abandoning (or at least pouring some cold water on) the hope for a science of utterance interpretation (as Fodor, Chomsky, Davidson, and others suggest).

So, cognitivists like Harris must take care to avoid excessive credulity about the prospects for the marriage between the internalist, Chomskyan theory of

semantic competence and a Fodor-like conception of semantic processing. As just reviewed, the existence of a Fodor-modular processing system for a particular domain is not entailed, logically or conceptually, by a Chomsky-modular competence theory for the information that is proprietary to the process in question.

That said, for the reasons noted above, it is generally agreed that the only source of evidence for a competence theory is performance data (Chomsky 1965: 4). Thus, the Fodorian conception of modular processing does suggest an attractive strategy: look to see if we have good reason to posit a Fodor module that deploys semantic competence (lexical and compositional knowledge) during the processing of linguistic stimuli. But this is at best a plausible expectation and an attractive research strategy; it is not an argument for the existence of a semantic Fodor module. And the attraction of this line of thought, of course, is that hearers do manage to solve Fodor's frame problem when it comes to understanding speakers, and so utterance interpretation is indeed a tractable process, carried out successfully by human minds. If the operations of a semantic, Fodor-like module, given a certain linguistic input, do not vary as a function of extra-modular influences, then this module affords hearers a reliable starting point for the complicated process of working out speaker meanings. The project of cognitivist semantics would then be to ascertain what this system outputs given linguistic input and how it generates those outputs.

Even if one agrees with Harris (as I am inclined to do) that an empirically adequate account of utterance interpretation is possible and also that such an

account must preserve the distinction between semantic and pragmatic information, it is not obvious that the only way to achieve that preservation is by appeal to Fodorian encapsulation. One can resist the inference from (i) the assumption that the semantics-pragmatics distinction carves psychological nature at a joint to (ii) the conclusion that the joint in question is an instance of the more general distinction between modular input-systems and non-modular cognition. After all, there are alternative cognitive architectures, architectures that are non-Fodorian (i.e., that reject encapsulation) but allow for individuation of cognitive processes in other ways.

For example, some have argued that many of what Fodor took to be central and hence non-modular processes are in fact modular (e.g., Sperber 1996; Carruthers 2006). For Fodor, very few cognitive processes are modular: the processes associated with the five sensory modalities, along with some linguistic processes (Fodor 1983), and perhaps one or two others, such as mindreading (Fodor 2000: 97). By contrast, proponents of a "massively modular" account of the mind argue for a much longer list of modular processes, as the label suggests. However, the list is longer precisely because massive modularists deny that Fodor-style encapsulation is constitutive of modularity.

In making the case for massive modularity, Peter Carruthers reconstructs the following Fodor-like argument for encapsulation of some cognitive processes. Any computational process must be tractable if it is to be carried out by a human mind: that is, a human mind must be able in principle to carry out the process in finite time. In order for the computational processes carried out by the human

mind to be tractable, they must be frugal: that is, both the amount of information required for the operations involved and the complexity of the algorithms used to process this information must be economical with regard to the task in question. Frugality, in other words, requires a solution to frame problems, since checking every hypothesis against every potentially relevant belief is not an economical strategy for belief fixation. If the mind is realized by a set of computational processes that are informationally encapsulated, this fact would begin to explain how frugality is achieved (Carruthers 2006: 14).

The Fodorian view of encapsulation has it that most of the information generally available in the mind cannot penetrate the operations of a modular system. However, as Carruthers argues, all that frugality requires is a kind of limited encapsulation. Frugality entails that the system's operations cannot be affected by most of the information generally available in the mind, but not because there is a determinate body of information that in principle is persistently blocked from affecting the system's operations. As Carruthers notes, the Fodorian understanding of strict encapsulation corresponds to one of two readings of the plausible but structurally ambiguous claim, 'A modular system's operations cannot be affected by most of the information generally available in the mind.' The Fodorian reads this with 'most' taking scope over 'cannot,' entailing that there is a determinate body of information that cannot penetrate the module. Alternatively, it could be that frugal operations cannot be influenced by most of the information generally available in the mind (with 'cannot' taking scope over

'most'); and this doesn't require any permanent isolation of the modular operations from a determinate body of information.

Limited but not strict encapsulation of cognitive processes, Carruthers argues, is supported by the literature on heuristic processing, which shows that frugality can be achieved without Fodorian encapsulation (Carruthers 2006: 16). Granted, Fodor's architecture is not incompatible with the use of some heuristics. And, of course, Carruthers' argument shows only that strict, Fodorian encapsulation is not entailed by considerations of tractability and frugality. Other, non-Fodorian notions of modularity that allow for some top-down, extra modular effects on (for example) visual processing might be sufficient for frugality. Hence, the choice among these notions of encapsulation is ultimately a matter for abduction: what notion of encapsulation (strict, Fodorian encapsulation or the limited encapsulation of massive modularists) makes for the best explanation of the relevant observations?

Although it eschews the strict, Fodorian notion of encapsulation, massive modularity does not thereby render us incapable of preserving the semantics-pragmatics distinction. We might instead attempt to functionally individuate semantic processes, just as one can functionally individuate visual processes (see Ogilvie and Carruthers 2016), and we can maintain that the semantic system consists of a proprietary set of mechanisms without insisting that it operates only on bottom-up input. At least in principle, this would allow us to locate semantics as a component of a serious psychological science of language, separated from pragmatics, without appeal to strict, Fodorian encapsulation.

For now, the lesson is that one could agree with Harris that an empirically adequate science of utterance interpretation must preserve the distinction between semantics and pragmatics and agree moreover that *some* notion of modularity will be crucial to its preservation, while rejecting the Fodor-like encapsulation hypothesis for semantics. The programmatic argument for encapsulation is therefore unsuccessful. However, whether this non-Fodorian conception of modularity in fact allows us to preserve a robust semantics-pragmatics distinction in a psychological science of utterance interpretation will depend on how much top-down penetration of semantic operations we observe. If meaning retrieval and composition turn out to be "shot through" with top-down pressures of pragmatic considerations, we may end up concluding that semantic operations are just as much a matter of belief fixation as inferences about speaker meaning. In chapter 3, I look at several pieces of evidence that suggest that semantic processing of linguistic input indeed appears to be regularly subject to rapidly detected pragmatic influences. And I show in chapter 3 how a proponent of the Fodor-like view might respond to these observations in a principled way. But I also ask (mostly in chapter 5), what would be left of the semantics-pragmatics distinction in a massively modular mind.

2.1.2 THE CLUSTER ARGUMENT

Another way in which the encapsulation hypothesis for semantics can be defended is by arguing that semantic operations exhibit the other characteristic features of Fodor modules. On the classic Fodorian account, central, non-modular processes exhibit a handful of properties: domain generality; sensitivity to the full

range of an agent's memories, beliefs, and other mental attitudes; accessibility, to some extent, to conscious awareness or introspection; and voluntariness, non-automaticity, and a relatively slow pace. Modular processes, on the other hand, also exhibit a handful of shared properties: domain specificity; two-way informational isolation, i.e., both encapsulation from and inaccessibility to central cognition; involuntariness, automaticity, and a relatively fast pace; a fixed neural structure; and characteristic patterns of breakdown and acquisition (Fodor 1983). Insofar as one agrees with Fodor that modules are a natural kind, unified by their possession of the characteristics in question, then observations that support attribution of a sufficient number of these features to a particular system might be taken to (inductively) support attribution of the rest.

Both Harris (2020: 5-10) and Borg (2004: 86-106) point out that semantic processing appears to have many of the properties of a Fodorian module, suggesting what I call 'the cluster argument.' It is observed, for example, that hearers quickly grasp the meanings of words they encounter and appear to do so without choosing to (provided that they are paying attention and are themselves speakers of the language in question). Furthermore, it is plausible that semantic processing is limited to a certain range of linguistic inputs. There is reason to think that semantic processing exhibits inaccessibility to other aspects of cognition that Fodor would call 'central': ordinary speakers do not have explicit beliefs about the principles of composition that are deployed during processing of speech, for example, or about the route they take from word identification to concept retrieval. So, one reason to think that semantic processing is

informationally encapsulated is the fact that it apparently exhibits many of the other features of Fodorian input systems.

There are two reasons why this argument is unsuccessful. First, setting aside the feature of informational isolation, one might reasonably question whether semantic operations exhibit all the other features that are constitutive of Fodor-modularity, as Harris and Borg both argue. For example, in saying that semantic operations are fast, what is the intended comparison class? Cognitive work that is undeniably post-semantic and not the result of encapsulated processing, such as determining the referent of a pronoun on an occasion of use, is "mind-bogglingly" fast (Fodor 1983: 61); data from visual world eye tracking studies indicate that the reference of a pronoun is resolved within 200ms of word onset (Arnold et al. 2000). During that same 200ms window, processes of syllable recognition must also take place. Since whatever semantic processes involved in pronoun resolution would presumably be carried out after syllable recognition, those semantic processes are presumably carried out even more rapidly than 200ms. And yet working out the referent of a pronoun on an occasion of use is a decidedly unencapsulated, highly inferential task. Hence, it would appear that being exceptionally fast is not unique to processes that are (by hypothesis) Fodor-modular; processes that are decidedly unencapsulated are also remarkably fast-paced.

Again, my claim here is only that one might reasonably question whether semantic operations are exceptionally fast-paced. Similarly, one can reasonably question whether semantic processing is supported by a stable neural structure

(see further discussion in sections 2.1.2 and 3.4), whether it is domain-specific, etc. In fact, I am inclined to think that semantic processing does exhibit some of the features constitutive of Fodor-modularity. However (and this is the second of the two reasons alluded to above), even if we grant that semantic operations exhibit some of the other features of Fodor-modules, this line of argument (like the programmatic argument) neglects alternative cognitive architectures that eschew encapsulation but preserve some of the other features that Fodor took to be jointly constitutive of modularity. On a massively modular account of the mind, there are modular systems that are realized by specific neural structures, and whose operations are both mandatory and inaccessible to higher-level cognition, but which are nonetheless *not* encapsulated in the strict sense Fodor has in mind (Carruthers 2006: 6-7). So, even if it is true that semantic operations exhibit a proper subset of the features of Fodor-modules, it does not follow that such operations are encapsulated. Semantic operations might be inaccessible to general cognition, mandatory, relatively speedy, and realized by a fixed neural architecture, but penetrable by higher-level cognition.

It is worth considering a further line of argument, one that is intended to support encapsulation not by way of induction from several features of Fodor-modules, but one such feature in particular. Fodor himself considers an argument for encapsulation of input-systems that rests largely on the observation that perceptual processes appear to be subserved by a stable neural structure, a distributed but nonetheless dedicated set of brain areas. In short, the argument has

it that encapsulation is a natural correlate of the existence of function-specific neural structures (Fodor 1983: 98-99).

Regardless of the merits of this kind of argument, in contrast with, e.g., vision and the distributed network of brain regions identified as its substrate (V1, etc.), no stable neural structure has yet been identified as the basis for semantic processing. Models have been proposed in the psycholinguistics literature. Some models locate the basis for conceptual-semantic processing in a distributed set of neural structures including the left posterior middle temporal cortex (PTC) and the anterior temporal cortex (ATC) (Lau et al. 2008). Other models point to the anterior temporal lobe (ATL) as the basis of semantic comprehension in particular (Matchin & Hickok 2020: 4-5). However, the sense in which 'semantic' is used in this literature is rather loose, sometimes referring to something quite broad, like deployment of world knowledge and conceptual associations, and sometimes referring to something more narrowly linguistic. In the sense in which I have used the term here, 'semantic' refers to something more narrowly linguistic, where the output of semantics given a word as input may or may not be conceptual. All parties, in any case, agree that not all conceptual activity that follows word recognition is directly caused by word recognition by itself; likewise, not all thoughts that follow recognition of a sentence are directly caused by recognition of the sentence by itself. Accordingly, in attempting to locate brain regions that do semantic work in the narrow linguistic sense of 'semantic,' we must remember that thoughts occur in a hearer's head independently of the unfolding speech, even when the hearer is an attentive one. That concepts are activated in a certain region

while speech occurs does not entail that these regions are activated as the direct causal result of narrow linguistic processing. (See more discussion of the neuroanatomy of semantic processing in chapter 3 below).

In any case, even if we were successful in identifying the dedicated neural basis for narrowly linguistic semantic operations, this would not entail the encapsulation of those operations. After all, massive modularists can and do agree that early visual processing is subserved by a dedicated, albeit spatially distributed set of neural structures (V1, V2, etc.; see, e.g., Ogilvie & Carruthers 2016: 729), but deny that the operations subserved by these structures are encapsulated from general cognition. Likewise, one might decide that some particular neural structure is the basis of the smallest semantic process—say, the left PTC—but deny that the operations of this system are encapsulated from top-down influences. Indeed, one could in principle find evidence to support both the conclusion (e.g.) that the PTC subserved the smallest semantic operation and the conclusion that activity in this brain area was modulated by extra-linguistic contextual features. Precisely this line of argument against the encapsulation hypothesis for vision is pursued in Ogilvie & Carruthers 2016, who draw on a range of experimental results. I discuss in more detail, in chapter 3 below, how this same strategy might be pursued regarding narrowly semantic processing.

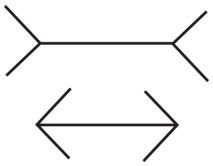
So, to conclude, the inductive argument from any subset of the features of Fodor-modules to the encapsulation hypothesis is rather weak. An unencapsulated processing system might have any number of the other features that Fodor took to be jointly constitutive of modular systems (relatively fast pace of operation,

inaccessibility to conscious introspection, a dedicated neural basis, characteristic patterns of development and breakdown) and still fail to be encapsulated in the sense of Fodor 1983. Hence, we should look elsewhere for considerations that might be more useful in assessing the encapsulation hypothesis.

2.1.3 THE POSITIVE EVIDENTIARY ARGUMENT

Yet another strategy is to point to empirical observations that are then argued to be best explained by appeal to the modularity hypothesis. Here, the argument for encapsulation of semantics mirrors well-known arguments for encapsulation of sensory processes, such as vision. However, while the relevant observations (mostly having to do with illusion-like phenomena) are indeed compatible with encapsulation, they are likewise compatible with unencapsulation. So, the illusion-based case in favor of encapsulation of semantics is deductively invalid and also inductively rather weak, as I argue here, building off work in the domain of vision.

A very common way to argue for the Fodor-style encapsulation of vision is by appeal to persistent visual illusions, such as the Müller-Lyer.[22]



The two horizontal lines are in fact equal in length, and yet they do not appear to be so, the line on the top appearing to be longer. However, even when you have

22 Originally reported in Müller-Lyer 1889.

used a ruler to convince yourself that the two lines are equal in length, they nevertheless continue to appear unequal. The observation of this kind of persistent visual illusion is often taken to suggest that early visual processing is informationally encapsulated: although one believes quite confidently that the two lines are equal, and despite the fact that this belief is highly salient and consciously available, this belief simply cannot influence our perceptual experience (Fodor 1983: 67; cf. Pylyshyn 1999). The large number of such persistent perceptual illusions is then pointed to in order to strengthen the case for Fodor-like modularity of sensory processing.

By analogy, one way to argue for encapsulation of semantic processes is to find evidence for something like a persistent semantic illusion. Harris (2020) attempts to do exactly this, borrowing a thought experiment from Pettit (2002). Suppose a hearer were led to have the false belief that, as a result of a procedure, he now hears mass nouns as having a meaning that they do not in fact have. Suppose the hearer then encounters a sentence like 'Let's have mud for lunch.' The hearer will believe that he has misunderstood the meaning of 'mud,' but the utterance will nonetheless seem to have the correct meaning to him. The intended lesson is that the hearer's false belief that he systematically misinterprets mass nouns has no effect on semantic perception. Following Pettit, Harris notes the analogy to the Müller-Lyer illusion: the fact that the illusory appearance of different lengths persists with the belief that the lines are equivalent is taken to be evidence for informational encapsulation of visual processing. Harris' thought experiment is meant to show that the same kind of argument can be given for

encapsulation of semantic processing, to include recovery of the meaning of a monomorphemic word (Harris 2020: 8).

This argument by analogy to visual illusions is unconvincing, however. Visual illusions do fit well with a strictly Fodorian conception of vision. But this is just to say that the Fodorian account of modularity, with its commitment to encapsulation, offers one way to explain the phenomenon. That said, explaining the phenomena in question does not require appeal to encapsulation. As Ogilvie and Carruthers (2016: 726) argue, the Müller-Lyer illusion shows that early visual processes over bottom-up input cannot be dominated by higher-level information; it does not show that such higher-level information cannot ever contribute to or influence early visual processes. And it is also possible that, in other circumstances, higher-level information might indeed dominate lower-level processing. Hence, as Ogilvie and Carruthers put it, when it comes to establishing Fodor-modularity, "the argument from illusions is simply invalid" (o.c.). The question comes down, again, to what the best explanation is for the relevant observations. And it is presumably in this abductive way that Fodor himself intended to argue for the strict encapsulation of visual processing based on persistent visual illusions. However, Ogilvie and Carruthers are correct to point out that the strict notion of encapsulation is not required to explain the observations related to visual illusions.

The same goes for the argument based on Harris' thought experiment for the Fodor-style modularity hypothesis for semantics. Even if one shares Harris' intuition that the hearer's beliefs cannot override the linguistic input in the 'mud'

thought experiment, this does not establish that higher-level cognition is incapable of having some influence on semantic operations over linguistic input. All that Harris' thought-experiment establishes is that the bottom-up linguistic input (sometimes) cannot be dominated by higher-level information, which is perfectly compatible with the unencapsulation (or limited encapsulation) of semantic operations. So, the argument from persistent semantic illusions to the Fodor-style encapsulation hypothesis for semantics is likewise invalid.

What is needed, then, is some evidence to suggest context-invariance in the mapping of a word-form onto a mental representation (one that might be called its meaning or value, as opposed to its form, on the one hand, or the totality of world knowledge about its denotation, on the other). As already indicated, polysemy therefore provides a case study for the encapsulation hypothesis, to be considered in detail in chapters 3-5. As I argue, if we are to find positive evidence for the encapsulation hypothesis, it will come in the form of evidence for something like stubborn activation of unmeant senses of a polysemous term.

2.1.4 THE NEGATIVE EVIDENTIARY ARGUMENT

In the other direction, one might argue for the encapsulation hypothesis by claiming the absence of any compelling evidence against it. Along these lines, Harris writes:

> If semantic composition were not an informationally encapsulated process—if it were sensitive to agents' beliefs—then we should expect strongly avowed false beliefs about semantic matters to have a deleterious effect on an agent's ability to create and understand meaningful utterances. But there is no evidence for top-down effects of this kind. Indeed,

> it would be truly shocking if we were to find evidence that believing certain semantic theories either enhanced or detracted from one's linguistic capacities, even a little bit. (Harris 2020: 9)

It is perhaps true that "there is no evidence for top-down effects of this kind," i.e., deleterious effects of confidently held but false beliefs about semantic *composition* on a hearer's ability to compose word meanings. However, the context-insensitivity of semantic *composition*, as opposed to meaning retrieval, is generally taken for granted. The assumption of Fodor-modularity for compositional processes therefore comes easily. Philosophical debates about semantic modularity have been focused on the lexical level, not the phrasal level. Subject to debate in philosophy of language is the question whether things like concept retrieval or the contribution of a word to direct speech act content are modulated by extra-linguistic context—not whether mode of composition is contextually modulated. In what follows, I set aside discussion as to whether the operations involved in semantic composition are encapsulated.

Though Harris does not discuss them, there are data that suggest that higher-level, extra-linguistic propositional attitudes do have top-down effects on just about every stage of utterance interpretation (see chapter 3). And, based on some of these data, practicing psycholinguists have indeed suggested that we cannot preserve a principled distinction between semantic and pragmatic processing in the study of language processing. For example, in reviewing the relevant body of experimental observations, Nieuwland & van Berkum conclude that "local [i.e., linguistic] and global [i.e., extra-linguistic] context can both immediately influence interpretation, without giving principled precedence to local semantic

information;" they then suggest that perhaps we cannot draw a principled distinction between semantics and pragmatics when it comes to the study of processing (Nieuwland & van Berkum 2006: 1108).

As a preview of the kind of data in question, consider the following prompt:

(9) How many animals of each kind did Moses take on the ark?

Most people respond to this question with the answer 'Two' (Erickson & Mattson 1981). The same effect is observed with a variety of similarly misleading sentences. Of course, on reflection, the correct response should be something like 'Don't you mean *Noah*?' However, most hearers believe that an utterance of (9) asks a question about Noah and answer accordingly. Is this evidence against Fodor-like encapsulation of the semantic operation of meaning retrieval? Probably not. It will depend on how we interpret this effect, and it is easy enough to imagine interpretations that avoid the assumption that what is going with (9) is semantic in the relevant sense (as opposed to, say, attentional and therefore pre-semantic, or inferential and therefore post-semantic). However, the so-called "Moses illusion" does suggest that we ought to avoid concluding too quickly that meaning retrieval is encapsulated. If one operation of the semantic system involves yielding singular concepts as output given proper names as input, then the Moses illusion might be interpreted as evidence that this operation is unencapsulated, being subject to top-down influences of world knowledge.

On the other hand, careful attention to effects like the Moses illusion might only serve to boost the prospects for the encapsulation hypothesis. That will depend again on how we interpret the observed experimental results, which

depends on considerations about experimental design, linking hypotheses, etc. In any case, experimental observations like this need to be addressed by proponents of encapsulated semantics, not ignored. To maintain that these observations provide no evidence for unencapsulation of lexical semantic processing (as, e.g., Harris must, in order for his 2020 case for encapsulated semantics to succeed), the proponent of Fodor-like modular semantics will have to offer explanations for the observations in question, explanations that preserve the encapsulation hypothesis.

In other words, one important job for the proponent of Fodor-modular semantics is to assess the experimental literature and explain away any apparent counterexamples in a principled fashion. And this is unsurprising. As noted above, with regard to perceptual and grammatical processing, Fodor spent quite a few pages showing how, by articulating and then appealing to a set of ground rules for analysis of putative evidence for top-down penetration, we can explain away many apparent counterexamples to the encapsulation hypothesis for a given domain (Fodor 1983: 72-85; on vision in particular, cf. Pylyshyn 1999 and more recently Firestone & Scholl 2016). Therefore, in the next chapter, I articulate those ground rules and then use them in assessing several experimental results that, at least initially, appear to be relevant to the encapsulation hypothesis for semantic processing.

Chapter 3: Is there evidence for unencapsulation of meaning retrieval?

As argued in the preceding chapter, debates about encapsulation for any given processing domain come down to a careful assessment of the available experimental record. Regarding vision, there are several experimental observations that suggest top-down, extra-modular effects. Proponents of Fodor-style modularity for visual processing seek to explain away the relevant observations on principled grounds. Skeptics of the Fodor-style view of modularity take issue with the explanations offered by Fodorians. The debate becomes a matter of inference to the best explanation for the observations in question. This is the same kind of dialectic that ought to be pursued, then, in the domain of semantic processing. The present chapter demonstrates as much, assessing various experimental observations that at least initially suggest that the process of content resolution (retrieving a concept given a monomorphemic word as input) is cognitively penetrable and therefore not modular in the Fodor-like sense. As I show, the proponent of the Fodor-like view can respond to this challenge. However, no defense of a Fodor-like view of semantics would be complete without some discussion of these data.

In his defense of Fodor-modular semantics, Harris (2020) says nothing about it, but there is indeed a large body of experimental literature that suggests top-down, context-driven influences on virtually every stage of the moment-to-moment processing of linguistic stimuli. For example, in reviewing a relevant

body of experimental observations, Nieuwland and van Berkum conclude that "local and global [i.e., linguistic and extra-linguistic] context can both immediately influence interpretation, without giving principled precedence to local semantic information;" indeed, these authors go on to suggest that we may wish to drop the idea that there is a sharp distinction between semantics and pragmatics altogether (Nieuwland & van Berkum 2006: 1108). If the Fodor-like hypothesis for semantics is to be maintained, the relevant experimental record must be evaluated and any observations that appear to be inconsistent with the encapsulation hypothesis must be explained away on principled grounds. Again, as argued above at the end of chapter 2, we cannot simply assume that the semantics-pragmatics distinction must be preserved in a science of language use, and then go about trying to locate a deep mental division that so preserves it.

In what follows, I point to several observations that initially appear to be inconsistent with the hypothesis that an encapsulated semantic processor maps a monomorphemic word to some uniquely determined mental entity. For example, consider again the Moses effect, discussed briefly at the end of chapter 2. In some contexts, it looks as if a speaker of 'Moses' is perceived as having meant NOAH, with hearers behaving as if they do not notice that the speaker uttered 'Moses' and not 'Noah' (Erickson & Mattson 1981). Also discussed below is what might be called the 'ham sandwich effect': it would appear that a speaker of 'the ham sandwich' is perceived, in some contexts, as having referred not to an especially salient sandwich but rather to a customer who has ordered one (Schumacher 2014;

see Nunberg 1979 for early discussion of this kind of deferred reference or predication).

Given such observations, should we conclude that meaning retrieval given a word-form as input is not effected by a Fodor-style module? No—at least, not too hastily. Regarding the Moses illusion, we do not know for sure whether hearers in fact perceive 'Moses,' without any error detection or conscious inference about plausible speaker meanings. In other words, the Moses illusion could be the result of top-down influences of attention on the input to a semantic processing system that is encapsulated. Top-down, context-driven effects of attention on the input to a processing system are compatible with the encapsulation of any processing that is over that input and internal to the system itself. Also, at least in principle, one might explain the Moses results by appeal to post-semantic inferential or pragmatic processing: we are generally cooperative as hearers, and therefore likely to accommodate all sorts of mistakes on the part of a speaker. And such post-modular inferential operations being unencapsulated does not in any way threaten the encapsulation hypothesis for semantic processing.

So, as I show in the present chapter, there are principled ways to explain the Moses effect and other relevant experimental data such that they are compatible with the encapsulation hypothesis for semantics. However, we simply do not know what the path is from (e.g.) the linguistic stimulus, 'How many of each type of animal did Moses take on the ark?' to the perception that the speaker meant NOAH. Likewise, by analogy, we do not yet know exactly why simultaneous visual presentation of human lips producing a 'ga' sound and auditory

presentation of a ‘ba’ sound result in a ‘da’ percept in the McGurk effect (McGurk & McDonald 1976). The McGurk effect can be explained in a way that is compatible with the encapsulation of sensory processing. However, the availability of such an explanation does not entail that it is correct. Cognitive scientists studying vision must weigh the available explanations against one another, as always. So it goes for semantic processing. The Moses illusion and the other observations that I marshal together in this chapter can be explained in a way that is compatible with the encapsulation of semantic processing. However, the availability of such encapsulation-friendly explanations does not entail that one of them is correct. The competing explanations must, as always, be weighed against both the evidence and one another.

In any case, the observations in question do cry out for explanation within a cognitivist framework. And here the contrast between the internalist (for whom meanings are mentally individuated) and the externalist (for whom meanings are individuated with respect to mind-external phenomena) becomes especially stark: the externalist does not care at all about questions about the processing of linguistic stimuli and so has no direct interest (qua externalist semanticist) in the Moses effect. At least from the cognitivist perspective described above (chapter 1), the internalist theory is answerable to such evidence.

3.1 Ground rules

Fodor’s classic 1983 defense of the encapsulation hypothesis for the visual modalities and some linguistic processes involves a handful of different

arguments. He appeals to general considerations about speed and computational tractability, as discussed in chapter 2. However, Fodor's case for encapsulation rests largely on his efforts to explain away apparent counterexamples. Fodor deploys a small number of ground rules for interpretation of experimental results that appear to suggest penetration of a cognitive process that has been hypothesized to be Fodor-modular. Armed with these ground rules, the Fodorian is able to defend the encapsulation hypothesis against many putative counterexamples. Here, I quickly review these ground rules, which will then be deployed in a discussion of the experimental record that is pertinent to the encapsulation hypothesis for semantics.

Fodor's first ground rule states that we must be mindful of the locus of the observed effect. It is true that, whatever perceptual systems yield as output, this must be reconciled with background information and the like; and this reconciliation might sometimes involve top-down "correction," as when one ultimately judges that the stick in the water is straight, despite its crooked appearance. However, in order to show that an input system is unencapsulated, one must show "that the locus of the top-down effect is internal to the input system" itself (Fodor 1983: 73-74). Pylyshyn elaborates on this ground rule, arguing (with regard to vision) that, when we heed it, all putative cases of penetration should be recognized as cases of one of the following: (i) intra-modular top-down effects, (ii) top-down effects on post-modular systems, or (iii) attentional effects that alter the input to the system, not the manner of its processing (Pylyshyn 1999). As noted by Ogilvie and Carruthers (2016), we

might add to this list a fourth category, (iv) McGurk-type, inter-modular effects. With these four considerations in mind, the Fodorian is well-fettled for defense against many potential counterexamples.

Fodor's second ground rule states that evidence for penetrability of some cognitive mechanisms that do work like that of input systems is not evidence that input systems are penetrable. Take, for example, the finding that relative cloze value (i.e., probability of a text completion relative to others) affects performance on various tasks, indicating that the mechanisms involved have access to expectations about what people will say. As Fodor argues, this shows only that some language-handling processes have access to the higher-level information (in this case, word frequency)—not that the narrowly linguistic input systems themselves have such access (1983: 74-75). By analogy, the fact that the broader interpretive system has access to high-level or extra-linguistic contextual information does not show that the narrowly linguistic-semantic component of that system has access to such information. In effect, this second ground rule is equivalent to Pylyshyn's second consequence of Fodor's first ground rule, namely, that in paying attention to the locus of a contextual effect we must consider the possibility of top-down effects on *post*-modular systems.

Fodor's third ground rule states that we need to carefully distinguish two issues: whether a given system is informationally encapsulated, and whether there is top-down information flow within said system. In effect, this also follows from Fodor's first ground rule, being equivalent to Pylyshyn's category (ii), intra-modular top-down effects. As an example, Fodor considers the phoneme

restoration effect. When a word-form is produced audibly but with one phoneme replaced by a non-linguistic sound, such as a cough, experimental participants report hearing the word in full, with a cough in the background (Warren 1970; Fodor 1983: 65). If such reports are reliable, they indicate that we hear (e.g.) 'legislature' with a cough in the background when the actual auditory stimulus is 'legi[cough]lature.' The overly hasty non-Fodorian might take the phoneme restoration effect to show that perceptual processing of a stimulus is influenced by the agent's beliefs and expectations, and is therefore not encapsulated (e.g., Bruner 1957; discussed in Fodor 1983: 66).

However, as Fodor argues, the phoneme restoration effect does not show that the language input system—one part of speech perception, among others—is penetrable. Fodor's strategy here, illustrated with the phoneme restoration effect but generalizable to many similar observations, is to argue as follows. If there is this kind of top-down, extra-modular feedback in a system, then the system in question is not a module; or (by way of denying the consequent), if it is a module, then we must explain away the appearance of feedback. Again, in general, the main task that awaits the committed Fodorian is to explain away misleading appearances. In order to show penetrability of the speech processing system, we would have to show that the information accessed is not specified at any level within the system. If the information accessed by the input system in the phoneme restoration effect is lexical information, as opposed to information about the speaker's beliefs or intentions, then the effect poses no threat to encapsulation (1983: 75-76).

If these ground rules are not followed, we are liable to conclude too quickly that a given cognitive process is unencapsulated. Any serious opposition to semantic encapsulation must avoid the pitfall of concluding too quickly that an observation shows top-down penetration of semantic operations as opposed to top-down effects on post-semantic operations; or that an observation shows top-down penetration of semantic operations as opposed to attentional effects on the input to said operations; or that an observation shows inter- as opposed to intra-modular top-down information flow. By heeding Fodor's ground rules, we can reconcile much of the pertinent experimental record with the encapsulation hypothesis for semantics; and we can do so in a principled, rather than ad hoc, fashion.

When appearances suggest that a preferred hypothesis is incorrect, its maintenance requires that we explain away the misleading appearances. At this point, no serious defense of the encapsulation hypothesis for (e.g.) vision could simply ignore the vast literature suggesting penetration of visual processing. The Fodorian's job is to explain away, in a principled fashion, the relevant observations, guided by the ground rules described above. And, indeed, this is exactly how the Fodorian hypothesis regarding vision is defended nowadays (again, see, e.g., Pylyshyn 1999; Firestone & Scholl 2016). So, one of the goals in what follows is to show how a parallel dialectic arises for semantics and pragmatics, and in ways that have gone underappreciated in the philosophical literature on this distinction, even among cognitivists. With regard to the cognitivist, this is not merely a complaint about missing citations or mistaken

priorities, since a cognitivist semantic theory is indeed answerable to the kind of experimental data I review here.

3.2 THE MOSES EFFECT

Speakers often make mistakes, and often enough these mistakes are detected by a hearer and either accommodated or pointed out (see Davidson 1986 for a rich discussion of this phenomenon). However, mistakes can also go apparently unnoticed by the hearer. For example, when asked 'How many of each type of animal did Moses take on the ark?', listeners have a tendency to respond 'Two,' even when they know that the story about the ark involves Noah, not Moses. (Though conventionally labelled a semantic 'illusion,' the tendency in the Moses effect persists only until the mistake is pointed out, at which point hearers become immune to it, in contrast with persistent visual illusions such as the Müller-Lyer.)

The Moses effect was first studied experimentally by Erickson & Mattson (1981), who showed that indeed hearers tend to respond to such "distorted" questions as if they were undistorted. The effect has since been widely replicated; see Park & Reder 2004 for a more recent overview, to which the present exposition is indebted. The effect is observed even in absence of time pressure, even when participants are told to look out for the relevant distortions, and even when participants are required to read the question aloud before answering it (Park & Reder 2004: 275).

What gives rise to the Moses effect? It is often described as a 'semantic' illusion. Here, however, the use of 'semantic' is the relatively loose one on which

‘Moses’ and ‘Noah’ are, in the operative sense, semantically related in virtue of conceptual associations between MOSES and NOAH (i.e., two different singular concepts that are associated with overlapping sets of other concepts, including, e.g., PATRIARCH and WATER). It is this relation of conceptual association that psycholinguists have in mind when they refer to the “semantic similarity” of ‘Moses’ and ‘Noah.’ Relatedly, not just any substitute for ‘Noah’ will give rise to the effect; e.g., hearers do not respond with ‘Two’ when the question asked is ‘How many of each type of animal did Nixon bring on the ark?’, as shown in the original study (Erickson & Mattson 1981). Whatever conceptual relations might obtain between ‘Moses’ and ‘Nixon,’ these are insufficient to give rise to the illusory effect. In addition to partial overlap between the conceptual associates of the distorted term and those of the correct term, the general fit between the distorted term and the linguistic context in which it is embedded also influences the emergence and size of the effect. This, too, is thought to be shown by the ‘Nixon’ example: it is both the relative lack of shared conceptual associations between ‘Noah’ and ‘Nixon’ and the poor fit between the latter and its immediate co-text that results in higher rates of error detection in the ‘Nixon’ conditions (van Oosterndorp & Kok 1990).

Does the Moses effect count as evidence against the hypothesis that a dedicated, encapsulated word processor translates a recognized word-form to some mental entity, such as a concept? At least at first glance, and given some ancillary assumptions, one might be inclined to think so. Suppose that one has reason to expect the outputs of a narrowly semantic word processor to be

concepts—i.e., mental representations like NOAH.[23] If so, then one interpretation of the Moses effect would have it that this processing system, given 'Moses' as input, sometimes yields as output MOSES, but sometimes yields instead NOAH. In other words, on one reading of these results, a speaker of 'Moses' can in some cases be perceived as having meant NOAH. This interpretation of the effect at least fits with the basic observation that hearers do not seem to detect the distortion in the 'Moses'-for-'Noah' condition: the hearer answers 'Two' because the hearer directly judges, without error detection or inferential cooperativeness, that the speaker has asked a question about Noah, because a processing system that receives 'Moses' as input yields NOAH as output. And the variation in the output given 'Moses' as input is driven by top-down pressures of co-text and world knowledge. So, on this line of thought, the Moses effect is an instance of top-down penetration of the word-level operations of the semantic processing system, operations which must therefore be unencapsulated.

That, in any case, is one line of thought. I shall not endorse it here. Instead, my goal here is to consider a variety of explanations for the Moses effect. If the Moses effect does not count as evidence against encapsulation of the word processor, this will be so because of the availability of a plausible explanation of the effect that is consistent with the Fodor-like modularity hypothesis for semantics. As with putative instances of top-down penetration of sensory

23 This is one of several hypotheses regarding what a dedicated semantic word processor might yield as output given a word as input. Instead of concepts, such a system might yield as output lists of concepts, or something altogether non-conceptual, such as a constraint, instruction, or rough indication of a region in conceptual space. These options are discussed in detail below in chapters 4 and 5. For now, I focus the discussion of encapsulation on a simple version of the hypothesis.

processing, the job of the cognitivist who endorses the encapsulation hypothesis for semantics is to explain away observations like the Moses effect, and to do so in a principled rather than ad hoc fashion.

Four different kinds of model have been offered in the experimental literature to explain the Moses effect (Park & Reder 2004: 278). Perhaps the simplest explanation is that listeners are generally cooperative: despite the behavioral evidence (i.e., the 'Two' answers), hearers or readers do notice the error (i.e., 'Moses'), but they are able to rapidly infer the speaker's intention (i.e., to ask a question about Noah), and so ignore or discount the error. A second explanation has it that the incorrect term is not actually processed—listeners fail to actually hear/read 'Moses' in the first place. On a third explanation, the incorrect term is correctly heard or read, but then the wrong informational or conceptual resource is retrieved from memory. Finally, a fourth explanation (closely related to the third) appeals to a partial match between the listener's representation of the question as a whole and the long-term memory structure that gives the answer.

The explanation in terms of cooperativity is consistent with the encapsulation hypothesis for word-level semantic processing. And, since we know that hearers often make cooperative adjustments "on the fly" (e.g., in cases of accidental malapropism), this explanation has the virtue of appealing to pragmatic processes that are already independently motivated. Indeed, the proponent of encapsulated semantics could suggest that it is precisely because the processing system context-invariantly maps 'Moses' to MOSES that listeners are

able to quickly detect the error, infer the correct intention, and then cooperatively issue a response to the inferred question instead of the question actually asked. In other words, the cooperative explanation for the Moses illusion heeds the first Fodorian ground rule for interpretation of experimental results that suggest penetration: be mindful of the locus of the effect. The cooperative explanation takes the effect to be post-semantic, and hence post-modular, being the result of general pragmatic inference (the work of central cognition, in the Fodorian architecture), downstream from semantic processing. As Fodor himself emphasizes, encapsulation is consistent with top-down effects on post-modular processing; after all, the outputs of encapsulated processing must be reconciled with expectations and beliefs generated by non-modular, central processes (Fodor 1983: 73).

However, we have good reason to reject this cooperative explanation for the Moses effect (Park & Reder 2004: 278). Again, the effect is observed even when participants are explicitly warned to look out for and report any such distortions in the questions asked; in fact, participants were observed to be better at detecting distortions when they were told to focus on the gist of the questions and ignore distortions (Reder & Kusbit 1991). So, when participants are told to be less cooperative, the illusion is stronger; and, when participants are told to be more cooperative, the illusion is weaker. Relatedly, the effect arises even when the task requires participants not to answer a question, but to verify the truth of a statement, a task that motivates less in the way of cooperation on the part of the listener or reader (Bredart & Modolo 1988; Oostendorp & de Mul 1990). Taken

together, these facts suggest that the cooperative explanation is incorrect: the Moses effect does not arise because listeners/readers detect but then ignore the incorrect term, inferring the correct intention.

The second explanation mentioned above is likewise compatible with the encapsulation hypothesis. If 'Moses' is not actually processed—not actually heard or read by participants—then the effect is not a matter of NOAH being retrieved given 'Moses' as input, since 'Moses' is never recognized in the first place, and so does not serve as input to any processing system. And this line of thought is plausible, at least initially. Hearers and readers do indeed make predictions about upcoming linguistic stimuli based on preceding text. It is possible that, when a particular word is strongly predicted in context, a hearer (or reader) might fail to encode it during processing. This recommends another explanation for the Moses effect that likewise heeds Fodor's general ground rule about minding the locus of an apparent effect of top-down intrusion. The idea would be that the Moses effect is an effect of attention on the input to the processing system: when a word is strongly predicted in a certain position in the sequence of input, we sometimes fail to properly attend to (and therefore fail to encode in the first place) part of the actual auditory or visual stimulus.

However, we have good reason to reject this explanation as well. In the original study, participants were required to first read the question aloud themselves before answering (Erickson & Mattson 1981). If the illusion is the result of an encoding failure, then we should see improved performance in detecting the incorrect term when participants produce the sentence aloud before

answering; however, we do not observe such improved performance. As Park and Reder discuss (2004: 279), one might attempt to salvage the encoding-failure explanation as follows: if the incorrect word is processed faster when the distortion is not detected, this could be taken to suggest that the distortion is not detected because the word is not recognized in the first place. In fact, however, reading times are slower in Moses-type experiments when the distortion is not detected (Reder & Kusbit 1991). So, the Moses effect is not due to overly hasty reading or failure to process or encode the "incorrect" term. Indeed, in most experiments that have recorded response times to Moses-type questions, there is a slowdown for distorted as compared to undistorted questions (e.g., Erickson & Mattson 1981). So, we have little evidence to support the hypothesis that the word 'Moses' is never recognized by any processing system, and lots of evidence to suggest that (at least, at some level of processing, albeit not consciously available) it is (Park & Reder 2004: 279-280).

Yet another explanation Park & Reder (2004) discuss has it that the stimulus (i.e., 'Moses') is processed and correctly encoded, but hearers/readers then retrieve the wrong concept or information from long-term memory. For example, it could be that the hearer retrieves information about the number of animals of each species that were on the ark, information that does not include anything about who brought them on board. As an alternative version of this explanation, it could be that 'Moses' is recognized and correctly processed, but then NOAH is retrieved and passed along on the way to an attribution of a speaker intention.

This kind of explanation for the Moses effect might be taken to suggest top-down penetration of the semantic word processor. Again, suppose one assumes that such a processor serves to retrieve a concept given a word as input. On this third account of the Moses effect, a hearer's mind can (at some level below conscious awareness) recognize 'Moses' but rapidly retrieve NOAH and pass the latter along on the way to an attribution of speaker meaning. And this happens with 'Moses' but not 'Nixon,' suggesting that the effect is an effect of background knowledge (i.e., knowledge of stories about biblical patriarchs). Thus, one might conclude that what underlies the Moses effect is top-down modulation of the operation of the semantic word processor.

However, this third explanation in terms of imperfect retrieval from memory is compatible with the encapsulation of the semantic processing system. Here is the rough line of thought. The effect is not one of attention on the input to the processing system. Rather, on this third explanation, the effect is arguably 'post-semantic.' That one ends up answering 'Two' does not show that the semantic system itself eventuates in activation of NOAH; it shows only that NOAH ends up being retrieved (i.e., activated above some threshold, to use another popular metaphor) and then passed along on the way to speaker meaning attribution. Again, suppose we have reason to think that a concept is the output of the semantic processor given a word as input. Then the encapsulation hypothesis predicts context-invariant mapping of 'Moses' as input to MOSES as output; the operations carried out over 'Moses' cannot be susceptible to pragmatic considerations (for example, background knowledge about biblical stories). And it

could be that a processing system does effect exactly that transformation during presentation of a Moses sentence, mapping 'Moses' to MOSES, although the latter plays no role in downstream processes involved in utterance interpretation, because top-down pressures result in greater activation of a closely associated concept. All the while, the hearer is unaware of the clash between the output of the semantic processor and the thought they end up entertaining about Noah, since the reconciliation is not brought about as a result of conscious inferential work. (A similar, encapsulation-friendly line of thought is available with regard to the fourth explanation for the illusion (i.e., in terms of partial match between sentence representation and queried memory structure).)

So, the explanations that are available for the Moses effect can be made to fit with a Fodor-like model of word-level semantic processing. The effect can be explained by appeal to the influence of attention on the input to the processor, or by appeal to top-down influences on post-modular processing. While the explanations in terms of post-semantic cooperative inference and pre-semantic effects of attention do not fit well with some of the available observations, all that is required by the encapsulation hypothesis is that the operations of the semantic processor, given a word as input, are insusceptible to the top-down influence of extra-modular information. On one version of this hypothesis, the semantic processor outputs concepts given monomorphemic words as input. Even if we grant that 'Moses' is detected (at some level) and reject the post-linguistic, inferential account of the effect, the fact that NOAH is sometimes retrieved from

long-term memory after hearing ‘Moses’ does not show that the semantic processor itself sometimes yields NOAH as output given ‘Moses’ as input.

So, one can indeed conclude that the Moses illusion does not count as compelling evidence against the encapsulation hypothesis for semantics. However, this conclusion is defensible only after considering the experimental record, under the guidance of Fodor’s ground rules. And here it is worth emphasizing the recourse to which the proponent of a Fodor-like modular semantics is left. We have no evidence that MOSES receives increased activation as a result of the presentation of the question about the ark, despite the fact that ‘Moses’ is part of said question. And yet it follows from the encapsulation hypothesis, together with a certain view of the semantic output given a word as input, that recognition of ‘Moses’ context-invariantly results in retrieval of MOSES, even when contextual considerations suggest the speaker is talking about Noah. So, while the experimental record concerning the Moses illusion does not falsify the encapsulation hypothesis, neither does it support said hypothesis.

Although the Moses effect does not end up being decisive with respect to the encapsulation issue, it affords some instructive lessons. A considerable amount of time elapses between onset of ‘Moses’ and the behavioral response of the experimental participant (i.e., the production of an answer). In between hearing the linguistic input and producing the response, enough time is available for quite a lot more than just narrowly linguistic processing to occur. Again, the experimental record fits poorly with the cooperativity hypothesis, on which the hearer detects the error in the question, then infers a plausible communicative

intention, and finally answers the inferred question about Noah. However, pragmatic operations need not be explicitly inferential or conscious in this way. Pragmatic operations can be carried out below the level of conscious availability, or 'sub-personally' (see, e.g., Recanati 2004). For example, again, the Moses effect could be the result of a non-linguistic process of spreading conceptual activation that occurs alongside and independently of the narrowly linguistic processing of the input. There is plenty of time between onset of 'Moses' and the production of the 'Two' answer for such a pragmatic process of spreading activation to result in higher activation of NOAH than MOSES, such that the hearer never consciously attends to the output of the narrowly semantic processor (i.e., MOSES), and so never attends to any thought (or question) about Moses.

As indicated above, this is merely an instance of the more general Fodorian strategy of defending encapsulation by locating any top-down effects outside the system in question. This might be taken to motivate the use of experimental measures that allow for more precise information about the timing of behavioral responses to linguistic stimuli. In fact, given the possibility of parallel (i.e., parallel semantic and pragmatic) processing, even a measure that is very fine-grained along the temporal dimension would not allow us to rule out the possibility that, alongside the operations of an encapsulated semantic processor that context-invariantly maps 'Moses' to MOSES, a non-linguistic process of spreading conceptual activation results in increased activation of NOAH as compared to MOSES (cf. Sikos et al. 2016 for a similar treatment of a related phenomenon). Still, consideration of more fine-grained measures should be

welcome, insofar as the goal is to appreciate the nature of the cognitive mechanisms that underlie the moment-by-moment processing of linguistic stimuli.

3.3 DATA FROM ELECTROENCEPHALOGRAPHY (EEG)

Electroencephalography (EEG) is an experimental method that provides temporally fine-grained data about the processing of specific stimuli by recording electrical activity on the scalp. The virtue of EEG recording is that it allows very fine-grained distinctions within a sequence of events, and therefore affords specific information about which stimuli cause which responses. Event-related potentials (ERPs) are brain responses to specific events identified via EEG recording. Temporally coarse-grained measures are not well suited to testing the encapsulation hypothesis for semantics, since they leave so much time for extra-modular processes to contribute to the experimentally elicited response. Hence, EEG perhaps offers a more promising experimental strategy for testing the encapsulation hypothesis. In what follows, I pursue this strategy.

One ERP that has been very widely discussed in the psycholinguistics literature is the N400 response, so named because it is a negative-going increase in electrical activity that begins about 200-300 milliseconds (ms) after stimulus onset, and then peaks about 400ms after stimulus onset. Most content words can elicit an N400 response (along with many non-linguistic stimuli—a qualification that becomes relevant to the discussion below). What is called the 'N400 effect' consists of differential modulation of the amplitude of the N400 deflection in

response to a pair of minimally different stimuli (see Lau et al. 2008 and Kutas & Federmeier 2011 for useful overviews).

Kutas & Hillyard (1980) first observed that an N400 effect results from comparison of sentence completions that are somehow anomalous or unexpected (10b) with sentence completions that are less anomalous or more predictable (10a).

(10a) I like my coffee with cream and sugar.

(10b) I like my coffee with cream and socks.[24]

The completion of (10b) with 'socks' gives rise to an N400 response with significantly greater amplitude than the N400 response that results from 'sugar' in (10a). This is the classic example of an N400 effect. The sentence-completion in (10b) is often described as 'semantically anomalous;' however, while it would be quite abnormal (and therefore anomalous) to put socks in coffee, this is not a semantic fact in the narrowly linguistic sense of 'semantic' that is under discussion here. (Likewise, the fact that humans and dogs tend to get along well is not a narrowly semantic fact about the word 'dog.') In any case, it would appear to be relative predictability given context, and not relative anomaly with regard to general background knowledge, that modulates processing difficulty. After all, an N400 effect is observed for non-anomalous but unpredictable continuations, as for example in a comparison of (10c) with (10a) (Kutas & Hillyard 1984).

(10a) I like my coffee with cream and sugar.

24 Here, I have followed the custom of underlining the 'target words,' i.e., the words that make for the minimal difference between the items, and whose visual or auditory onset marks the beginning of the event to which the N400 response is linked.

(10c) I like my coffee with cream and honey.

In related work, it was shown that words that are "semantically" primed by preceding linguistic input (as when an isolated presentation of 'tea' follows an isolated presentation of 'coffee') result in lower-amplitude N400 responses than words that are not primed (as when 'tea' follows 'chair') (Rugg 1985). N400 effects in this "semantic" priming paradigm tend to be smaller than in the sentence completion paradigm; however, the latency and spatial distribution of the N400 effects in these two paradigms are similar enough to convincingly indicate a shared underlying response (Kutas 1993; Lau et al. 2008: 921).

On the assumption that increased electrical activity reflects increased processing cost, the N400 effect reflects a difference in the processing cost induced by a content word given its fit with preceding context. Hence, this effect "seems likely to provide useful information about the timing, classification, and interactions of cognitive processes involved in natural language comprehension" (Kutas & Hillyard 1980: 204). In particular, the hope has been that the N400 (and ERPs more generally) would allow us to ascertain when different sources of information (both linguistic and non-linguistic) make their contributions to utterance interpretation as it unfolds moment-by-moment (o.c.). This line of research in psycholinguistics thus appears to be promising with regard to the encapsulation hypothesis, since whether this hypothesis is correct is largely a matter of whether there is any stage of sentence comprehension at which semantic operations are carried out independently of any top-down pressures of context.

And, indeed, the general conclusion among psycholinguists is that the experimental record regarding the N400 shows rapid, top-down effects of context on very early stages of sentence processing. As shown in the seminal study, co-text (or linguistic context) plays a role in generating the N400 effect, a finding that has been frequently confirmed in similarly designed follow up experiments. For example, in a story about how quick Joe was to get ready, 'slow' in (11a) gives rise to an increased N400 response in comparison to 'quick' (11b) (van Berkum et al. 1999).

(11a) Jane told Joe he was really slow.

(11b) Jane told Joe he was really quick.

As van Berkum (2008: 377) writes in a review paper focused largely on the N400, "one thing we observe again and again in ERPs is that the brain very rapidly relates the words of an unfolding sentence to the wider discourse."[25] Importantly, linguistic context (i.e., co-text) is not the only contextual factor that influences the amplitude of the N400 response to a word. For example, it has been observed that standard features of the extra-linguistic context, like who is speaking (or facts about who is speaking), can modulate N400 amplitude. When auditorily presented with the voice of a young child, 'wine' in (12) gives rise to a significantly increased N400 response as compared to presentation of the same sentence with an adult voice (van Berkum et al. 2008).

(12) Every evening I drink some wine before I go to sleep.

25 Neuroimaging studies give rise to a similar conclusion: "high-level, real-world knowledge can have an immediate effect on language comprehension" (Nieuwland & van Berkum 2006: 1098, citing Hagoort et al. 2004, Münte et al. 1998). See further discussion below in section 3.4.1.

Basic world knowledge (i.e., that young children do not usually drink wine) has a very rapid effect on sentence processing—an effect that is detected in electrophysiological recordings within just a few hundred milliseconds of word onset. The same conclusion is reinforced by similar studies: contextual influences on linguistic processing are detected very rapidly (see useful overviews in Schumacher 2014 and van Berkum 2008).

Does the experimental record concerning the N400 *prove* that the semantic processor is not encapsulated, but rather subject to the top-down influences of extra-linguistic context? No—but it does at least *suggest* unencapsulation. The relevant results do establish that extra-linguistic contextual information influences very early stages of sentence comprehension, with the effect detected within a few hundred milliseconds of word onset. And some of the authors of these experiments have endorsed much stronger conclusions from their N400 observations, including what amounts to rejection of the encapsulation hypothesis for semantic processing. As discussed below, for example, Nieuwland & van Berkum conclude from one such N400 study that extra-linguistic contextual factors can influence the earliest stages of semantic processing in such a way that, from the perspective of the psychology underlying language use, perhaps we cannot draw a principled distinction between semantics and pragmatics after all (Nieuwland & van Berkum 2006: 1108).

However, as I argue, we should not conclude too quickly from the EEG data that the encapsulation hypothesis for semantic processing is false. Instead, what is needed here is a careful analysis of the relevant results, guided not only by the

theories that have arisen in the psycholinguistic literature on the N400, but also by the Fodorian ground rules specified above (section 3.1). After reviewing in more detail the theoretical discussion of the N400 that has arisen in the literature, I focus on two case studies in particular, both of which appear to show very early top-down effects of extra-linguistic context on word processing.

3.3.1 INTERPRETING THE N400

Broadly, two kinds of interpretation of the N400 effect have been defended. On the original interpretation, the differential amplitude of the N400 response reflects processes involved in the integration of the critical word ('socks' versus 'cream') with the unfolding co-text (Kutas & Hillyard 1980; cf. Hagoort 2008, Oosterhout & Holcomb 1992; Brown & Hagoort 1993). On an alternative interpretation, the N400 effect is a reflection, not of integration processes, but contextually driven facilitation of (features of) long-term memory representations associated with lexical items (Kutas & Federmeier 2000; Federmeier 2007). Following Lau et al. 2008, I call these the 'integration view' and the 'lexical view,' respectively.

The integration view rests largely on the plausible hypothesis that integration of a word into a mental representation of an unfolding sentence is more difficult when the word is incongruent with preceding co-text, and therefore less predictable than congruent continuations.[26] An early observation that appears to support the integration view comes by way of priming studies. No differential

26 The same goes for integration of the word's meaning into representations of either the meaning of the unfolding sentence or the representation of the broader discourse. The relata for the posited integration relation are not always made clear and seem to vary across different versions of the integration view (Lau et al. 2008: 920, 923).

amplitude of the N400 response (i.e., no N400 effect) is observed for 'tea' when the priming word ('coffee'/'chair') is 'masked,' or presented too quickly to be consciously detected (Brown & Hagoort 1993; Lau et al. 2008: 921). It has been argued that this observation favors the integration view over the lexical view. Although masking prevents the hearer or reader from being consciously aware of the priming word, it is generally thought that the processes of lexical access are not consciously controlled but triggered automatically by certain bottom-up stimuli. On this assumption, the lexical view predicts an N400 effect in the masked priming conditions; but no such N400 effect was observed in early masked priming studies. Hence, masked priming results were taken to support the integration view. Additionally, some have argued that the N400 occurs too late to reflect lexical access, and so must reflect a post-access process, such as integration (Sereno et al. 1998; Hauk & Pulvermuller 2004; Hauk et al. 2006). These considerations, taken together, provide some support for the integration view of the N400 effect.

On the lexical view, the N400 effect arises when one of the two words being compared is more predictable in the linguistic context than the other, such that the more predictable word (along with its features) is therefore easier to access (i.e., it is easier to access the word's representation in long term memory, perhaps together with certain features tied to it). As the lexical view predicts, the very factors that are known to influence lexical access also influence the amplitude of the N400 (Lau et al. 2008: 921): word frequency (van Petten & Kutas 1990; Allen et al. 2003), repetition (Rugg 1985), and genuine versus pseudo-wordhood

(Holcomb & Neville 1990; Holcomb 1993). In addition, contrary to earlier findings (Brown & Hagoort 1993) more recent studies that used masked priming have shown an N400 effect when the interval between masked prime and target word is short enough (Kiefer 2002; Grossi 2006; Lau et al. 2008: 921-2). Hence, there are also considerations that favor the lexical view.

However, as Lau and her coauthors note, the factors that influence ease of lexical access might also influence ease of post-access integration (2008: 922), such that much of the experimental work on the N400 fails to adjudicate between the lexical and integration views. Accordingly, Lau and colleagues focus on the brain regions that are specialized for different aspects of semantic processing, and then use available data concerning the localization of the N400 response to figure out what specific operations (access or post-access) underlie the effect. Using this strategy, Lau and colleagues argue that the available data "strongly suggest that at least some substantial part of the N400 effect reflects facilitated lexical access, and thus that the N400 cannot be attributed only to post-access processes" (Lau et al. 2008: 928).

Clearly, the lexical view is the one that has the greatest potential to threaten the encapsulation hypothesis for semantic processing. First, the lexical view is not necessarily limited to the hypothesis that the N400 reflects the ease with which one accesses the representation of a word by itself (i.e., the ease of word recognition); the lexical view allows for the possibility that the access-level processes underlying the effect also include activation of some of the features that are narrowly associated with the word in long-term memory (Lau et al. 2008:

921), possibly to include features that place restrictions on concept retrieval. On the integration view, the N400 effect reflects processes that are downstream of the kind of narrowly linguistic, word-level semantic process that is my focus. However, on the lexical view, the N400 effect might be taken to reflect facilitation of precisely those lexical-level semantic processes (in addition to mere word recognition) that are pertinent to the encapsulation hypothesis that is my focus here. Accordingly, if observations concerning the N400 effect are to have any bearing on the encapsulation hypothesis for lexical semantic processing, this would require endorsement of the lexical view. Note that the case studies reviewed here are not meant to settle the dispute between proponents of the lexical and integration view. Rather, I am taking for granted the lexical view and then asking whether, on this view, some of the especially pertinent N400 data might meet the high standards for a counterexample to the encapsulation hypothesis.

3.3.2 AMOROUS PEANUTS

Whether "out of the blue," or in co-text in which an ordinary peanut is described as having properties that are typical of ordinary peanuts, 'in love' in (13b) would typically give rise to an N400 response with a greater amplitude than that following 'salted' in (13a) (Nieuwland & van Berkum 2006).

(13a) The peanut was salted.

(13b) The peanut was in love.

This is a standard N400 effect, since ‘in love’ is less predictable in the sentential context than ‘salted,’ and so (depending on the view one holds) either harder to access from memory, or to integrate in the unfolding sentential, semantic, and/or discourse representation. However, when read after co-text that describes an anthropomorphized cartoon peanut who has been dancing and enjoying a night out with another cartoon legume, the N400 effect is reversed, with ‘salted’ in (13a) giving rise to a greater N400 response than ‘in love’ in (13b) (o.c.).

Nieuwland and van Berkum take this observation to show that what they call a ‘one-step’ model of processing is favorable to a ‘two-step’ model. On a two-step model, hearers first locally compute the context-independent meaning of phrases and then integrate the results into a representation of prior discourse. Nieuwland and van Berkum cite, as proponents of such a two-step model, classic proponents of the encapsulation hypothesis for linguistic processing in general (Forster 1979; Fodor 1983; Kintsch 1988). By contrast, on a single-step model, semantic evidence supplied locally (i.e., via linguistic co-text) has neither temporal nor functional precedence over globally (i.e., extra-linguistically) supplied, pragmatic evidence. As proponents of such a one-step model, Nieuwland and van Berkum cite opponents of the Fodor-like, encapsulation-based model of linguistic processing (e.g., Trueswell and Tanenhaus 1994; Nieuwland & van Berkum 2006: 1099).

The single-step account directly predicts that in certain contexts (13b) should be very easy to process, since both semantic and pragmatic constraints are capable of overriding one another. The two-step model does not generate the

same prediction, since semantic constraints are given functional (if not temporal) precedence over pragmatic constraints. (Of course, as I discuss shortly, the two-step model (which involves a commitment to Fodor-like encapsulation) could in principle be made to fit with the relevant observation by appeal to other considerations, e.g., the ground rules described above in section 3.1.) According to the authors, the N400 effect in the cartoon condition shows that, as predicted by the one-step model, local and global factors can both immediately influence interpretation; local semantic information is not given any functional precedence in processing. Instead, extra-linguistic contextual factors can influence "initial interpretation of unfolding input"—i.e., the earliest stages of semantic processing. Indeed, the authors go so far as to explicitly say that, after all, perhaps we cannot draw a principled distinction between semantics, as knowledge of core word meanings, and pragmatics, as knowledge about word usage (Nieuwland & van Berkum 2006: 1108, citing as precedent Gibbs 1984; Clark 1996; Kempson 2001; Jackendoff 2002).

As argued above in chapter 2, this would be an overly hasty conclusion, since a semantics-pragmatics distinction at the level of competence does not entail the existence of a correlated division in language use. Again, the fact that knowledge of the sounds of birds is distinct from knowledge of their shapes does not entail that my brain includes two isolated processing systems, each of which makes use of only one of the two stored knowledges. In the other direction, evidence that a single processing system makes hybrid use of these two kinds of information in order to issue judgments about bird species does not entail that the

difference between knowledge of sound and knowledge of shape (never mind the difference between sound and shape) is a theoretical artifact. Likewise, the existence of a processing system that draws on linguistic and non-linguistic sources of information in parallel does not entail that there is no distinction between knowledge of language and knowledge of other things.

However, the peanut study does show that, at best, the encapsulation hypothesis for semantic processing sits uncomfortably with the available experimental record, Moreover, psycholinguists have occasionally suggested (as Nieuwland & van Berkum do) that the hypothesis is false. According to the encapsulation hypothesis, the operations of the semantic processor over linguistic input are insusceptible to any top-down, extra-modular influences of pragmatic considerations. Whatever mental entity 'peanut' is context-invariantly mapped to, it presumably does not have the feature +ANIMATE. And yet, in the right conversational context—i.e., in the course of a lengthy narrative about an anthropomorphized peanut—readers' brains rapidly reflect processing difficulty in response to predicates that clash with the presumption that the referent of 'peanut' is animate. Whether or not this is inconsistent with the encapsulation hypothesis is discussed below. For now, all I wish to point out is that a reasonable enough reaction to this kind of study would be to suspect that (as Nieuwland and van Berkum suggest) there is no interpretive stage during sentence processing that is both semantic (in the sense that it mediates somehow between syntax and thought) and encapsulated.

More generally, it has been widely argued that most functionally pre-

semantic processes are susceptible to some degree of dop-down influences of extra-linguistic context, to include syntactic parsing (Tyler and Marslen-Wilson 1977; MacDonald et al. 1994; McRae et al. 1998; Trueswell et al. 1994; for review, see van Gompel & Pickering 2007) and word recognition (see an especially convincing meta-analysis in Lucas 1999; see also chapter 1 above). Whether or not these arguments are successful, and whether parsing and word recognition are in fact unencapsulated, is of course subject to continued debates, not to be settled here. However, a growing consensus in psycholinguistics seems to favor models in which linguistic processes are unencapsulated. Hence, from the perspective of practicing psycholinguists, it is unclear why we would expect that semantic processing should be an exception in this regard.

To reiterate once again, my focus is what I have called the "smallest" semantic process, whereby some kind of composable mental entity is retrieved given recognition of a word. My point here is only that recent trends among practicing psycholinguists suggest that we should expect to find that this very small semantic process is unencapsulated, subject to top-down influences of extra-linguistic context, just as are the neighboring processes of parsing and word recognition (never mind functionally downstream processes, such as inferring implicatures, that are undeniably pragmatic and therefore clearly not Fodor-modular). My claim here is only that, based on what appears to be the general consensus among psycholinguists, a consensus that emerges from a long record of controlled experiments, we should expect a semantic word processor to be encapsulated only in the limited sense of (e.g.) Carruthers 2006, not in the stricter

sense of Fodor 1983. And, as we see already in a brief tour of the N400 effect and the peanut-in-love case study, some experimental evidence appears to meet this expectation, showing top-down influences on early semantic processes. Perhaps this appearance is misleading (see below, section 3.3.4). However, if so, then the encapsulation proponent owes some explanation for why we have been misled here, just as (e.g.) the encapsulation proponent with regard to auditory processing owes an explanation for the McGurk effect.

3.3.3 IMPATIENT SANDWICHES

Another example of an apparent contextual effect on something like meaning retrieval comes from studies of the moment-by-moment processing of deferred reference (Schumacher 2011, 2014). Deferred reference is the phenomenon whereby a term that would typically be used to refer to one kind of thing (e.g., 'ham sandwich') is instead used to refer to something that is not the typical referent, but related to it by way of background knowledge, or other pragmatic considerations (e.g., in the way that a sandwich is related to someone who has ordered or eaten one) (see Nunberg 1979 for an early and influential discussion of this phenomenon).

(14) The ham sandwich wants to pay.

In certain contexts, a speaker of (14) can be readily understood to have said that a salient customer who ordered a ham sandwich wants to pay his bill.

In her 2014, Petra Schumacher contrasts this observation with one of Nieuwland and Van Berkum's—namely, the judged unacceptability of (13a) as

compared to (13b), as reflected in the N400 effect, in a discourse about a cartoon peanut enjoying a date with another cartoon legume.

(13a) The peanut was salted.

(13b) The peanut was in love.

According to Schumacher, these two observations—the peanut in love and the impatient sandwich—together show the complex "interplay of linguistic meaning, context and truth-conditional content" during language processing (Schumacher 2014: 152). Accordingly, one goal for psycholinguists is to "disentangle the time-course of utterance interpretation as it unfolds in a moment to moment manner" (o.c.). As indicated above, one might hope that, in pursuing such a disentanglement, we will find some timing data that are pertinent to the encapsulation hypothesis. From the perspective of the moment-by-moment processing of linguistic stimuli, Schumacher defends two main conclusions: first, that context has an early influence on every stage of processing; second, that certain theoretical distinctions end up getting set aside when we attend to actual moment-by-moment processing (2014: 153). At least on a first glance, these experimental results appear to be relevant to the encapsulation hypothesis for semantic processing.

Schumacher (2011, 2014) supports this conclusion with a series of EEG studies focused on 'ham sandwich'-like sentences. In Schumacher's 2011 experiment, sentences like (15a) and (15b) were presented visually.

(15a) The doctor asks his assistant again who had called that early. The assistant responds that the therapist had called that early.

(15b) The doctor asks his assistant again who had called that early. The assistant responds that the hepatitis had called that early.

Given the question that is raised by the first sentence (who called?), the initial syntactic structure suggests that 'responds that' will be followed by a term referring to a potential agent of an event of calling or telephoning, as in (15a). So, it would not be surprising if we were to find that readers are more surprised by 'the hepatitis' in (15b) than they are by 'the therapist' in (15a), at least briefly. Therapists, but not diseases, are potential agents of telephoning events. And so, it would not be surprising to find an N400 effect when comparing responses to (15a) and (15b). However, no such N400 effect was observed in the experiment reported in Schumacher 2011. According to Schumacher, the absence of an N400 effect here is explained by the fact that the first stage of sentence processing is sensitive to extra-linguistic context (2014: 162) and is therefore an unencapsulated process. On Schumacher's view, the medical context that is established by the first sentence, together with world knowledge about the relation between medicine and disease, predictively facilitates parsing of the target sentence in (15b) just as much as it facilitates parsing of the target sentence in (15a).

In her 2011 experiment, Schumacher detected what is called a 'Late Positivity.' A Late Positivity is an ERP that manifests as a positive-going waveform that peaks around 600ms after stimulus onset (hence, it is sometimes called a 'P600' response). It has been argued that the amplitude of the Late Positivity is a reflection of processing costs incurred during integration of

linguistic meaning with a broader representation of the unfolding discourse. For example, the amplitude of the Late Positivity in response to 'the pistol' in (16) is lowest when the preceding text includes 'shot,' higher when the preceding text includes 'killed,' and higher still following 'found dead' (Burkhardt 2007).

(16) Yesterday, a PhD student was [shot / killed / found dead] downtown. The press reported that the pistol was probably from army stocks.

This suggests that the amplitude of the Late Positivity reflects difficulty involved in establishing discourse coherence: when co-text includes mention of a shooting, establishing discourse coherence in response to 'the pistol' is less taxing than it would be if co-text included no specific mention of a shooting event. Based on this interpretation, Schumacher hypothesizes that the Late Positivity that arises for 'the hepatitis' in (15b) as compared to 'the therapist' in (15a) is a result of the fact that her posited second stage of sentence processing still has to do some extra work to establish discourse coherence upon encountering the target word in (15b).

In her 2014 follow up, Schumacher reports that, by contrast, both an N400 effect and a Late Positivity are observed when the same target words are presented in sentential contexts that do not establish a context that supports the meaning transfer—e.g., by replacing 'the doctor' in (15a, b) with a proper name, as in (15a', b') (Schumacher 2014).

(15a') Thomas asks Claudia again who had called that early. Claudia responds that the therapist had called that early.

(15b') Thomas asks Claudia again who had called that early. Claudia responds that the hepatitis had called that early.

In a comparison of the ERP responses to the underlined target words in (15a') and (15b'), both an N400 effect and a Late Positivity were observed in Schumacher 2014, where the co-text includes proper names that are not supportive of the meaning transfer required for correct interpretation of the use of 'the hepatitis.' But only the Late Positivity, with no N400 effect, was observed in Schumacher 2011, where the co-text was contextually supportive of the transfer. Here is how Schumacher explains the differing results within her two-stage model of sentence processing. In the first stage, a parsing mechanism incrementally computes predictions for upcoming words, relying on certain aspects of context—both linguistic context (i.e., intra- and extra-sentential co-text) and extra-linguistic context. In the second stage, reasoning and inferential capacities are engaged in order to arrive at a meaningful interpretation, often via a process of pragmatic enrichment.

Crucially, both stages are influenced by contextual factors, and these factors go beyond linguistic context. As Schumacher states, "to account for the data in a proper way, a broad notion of context must be adopted," where the context in question includes "sentential context, co-text, situational context, world knowledge, common ground, etc." (2014: 152). So, although Schumacher's appeal to two stages of processing may seem to clash with Nieuwland and van Berkum's (2006) argument for a single stage of processing, this is not the case. In rejecting what they call a two-step model, what Nieuwland and van Berkum deny is the existence of a context-blind initial stage of semantic processing. And Schumacher would agree that there is no privileged stage at which only the

linguistic input influences semantic processing. Instead, contextual effects are predicted at every stage, though some contextually influenced processes are functionally (and temporally) downstream from others (e.g., Schumacher's stage two processes as compared to stage one processes).

When preceding sentential co-text does not support the reference transfer (as in the 2014 experiment, with proper names replacing content words that are supportive of the transfer), there is a greater processing cost as a result of a greater mismatch between what is predicted by the stage one parsing mechanism and the actual input. When preceding sentential co-text does support the reference transfer (as in the 2011 experiment, where preceding co-text includes content words that are supportive of the transfer), there is no observed difference in the processing cost across the two conditions, since there is no such mismatch between what is predicted by the stage one parsing mechanism and the actual input. Again, the stage one parsing mechanism in Schumacher's model is sensitive to extra-linguistic context, and so 'hepatitis' is not unexpected in the context of (15b).

The observed Late Positivity in both the 2011 and 2014 experiments is explained as follows. In the 2011 study, an increased Late Positivity is observed following 'hepatitis' in (15b) because some extra inferential work is required in order to build a representation of speaker meaning (and a broader representation of the discourse) as compared to (15a) (Schumacher 2014: 158). In the 2014 study, although the preceding co-text does not support the meaning transfer, the Late Positivity following 'hepatitis' shows that the transfer occurs nonetheless: the hearer still engages in stage two inferential processing that is required in order

to attribute a plausible speaker meaning and establish discourse coherence (Schumacher 2014: 162).

None of the processing models in the studies reviewed here explicitly includes an isolated stage of semantic processing at which a composable mental entity is retrieved given recognition of a disambiguated word-form. It is left for one to assume that some component of these models includes this smallest semantic process as a subcomponent. So, at least in principle, in denying that the stages in their respective models are encapsulated, the psycholinguists discussed here may not be committing themselves to denying that this very small stage of semantic processing (going from word to composable mental entity) is encapsulated. Perhaps the smallest semantic process is an exception to what seems to be the general rule regarding top-down penetrability in linguistic processing. However, one possible take-away here would be that concept retrieval is not an automatic, informationally encapsulated process: given sufficient contextual support for a concept (e.g., CUSTOMER) only indirectly related to the lexically encoded meaning of a term (e.g., SANDWICH), the semantic system can directly retrieve the indirectly related concept, without accessing the lexically encoded concept. Schumacher does seem to draw a similar conclusion (one she takes to be consistent with other EEG findings):

> …as a sentence unfolds and a partial interpretation of the utterance has already been formed, the space of alternative continuations is narrowed down based not only on linguistic knowledge, but also on world knowledge and common ground among other things. This supports models that assume an early influence of context on propositional content and suggests that the initial interpretation step is not merely confined to locutionary content (Grice's original conception of "what is

> said") but relies on context-dependent (pragmatic) information as well (as for instance suggested by Relevance Theory or truth-conditional pragmatics). (Schumacher 2014: 156)

In other words, the ham sandwich case reinforces the lesson of the peanut case: there is some experimental evidence that suggests top-down penetration of very early semantic processes, and many practicing psycholinguists have concluded on the basis of this evidence that, indeed, there is no stage of semantic processing that is encapsulated in the sense of Fodor 1983. Instead, the semantic processor, if it is encapsulated in any way, must be so in the limited sense of massive modularists, such as Carruthers (2006), as described above in chapters 1 and 2 (see further discussion also in chapter 5).

Note that this is not a case in which a hearer's consciously held beliefs influence concept retrieval, and so differs from Harris' 'mud' thought experiment (see chapter 2 above) in an important respect. The intuition we are supposed to share in response to the 'mud' example is that the speaker's explicitly held belief (i.e., that he systematically mishears the meanings of mass nouns) does not influence semantic perception. In the 'ham sandwich' experiments, by contrast, it appears that largely unconscious conceptual associations influence what the semantic system retrieves given a word as input (at least, on one interpretation of the results). In the right context, a hearer of (14) does not consciously attend to the premise that people often eat ham sandwiches, or the premise that ham sandwiches do not have propositional attitudes, on the way to judging that the speaker is asserting that a person (not a sandwich) wants to pay their bill. Nonetheless, if an indirectly related concept can be retrieved by the semantic

system, given a certain word as input, and if this variation in retrieval is a result of top-down pressures of context (e.g., background knowledge), this would falsify the Fodor-style encapsulation hypothesis for semantics—at least, on the assumption that concepts are the output of what I have called the smallest semantic process.

3.3.4 Defending encapsulation

I now discuss how those committed to a Fodor-like process of concept retrieval might explain the observations reviewed above. However, it is necessary to briefly and directly articulate an interpretation of these observations on which they threaten the encapsulation hypothesis, so as to clarify the nature of the challenge. First, the EEG literature concerning the N400 effect broadly suggests that top-down influences arise at just about every stage of utterance interpretation (van Berkum 2008; Schumacher 2014). This suggestion is in keeping with results emerging from neighboring literatures that deploy experimental measures other than EEG; for example, a similar conclusion is derived from fMRI studies that test the role of extra-linguistic context in early linguistic processing (e.g. Hagoort et al. 2004; Münte et al. 1998; cited in Nieuwland & van Berkum 2006: 1098). The encapsulation hypothesis that is my target here concerns a very "small," and very narrowly linguistic process, one that often is left implicit in psycholinguistic models of sentence processing—namely, translating a recognized word-form into a composable mental entity (for example, a concept). These N400 findings do not directly threaten this hypothesis. However, to maintain the encapsulation

hypothesis for word-level semantics, when just about every neighboring process exhibits top-down influences of extra-linguistic context, requires special pleading.

Second, there is at least one framework on which the N400 effect can be taken to reflect the very sort of narrowly semantic processing that is under discussion here in connection with the encapsulation hypothesis. On the lexical view of the N400, its amplitude is a reflection of the difficulty involved in accessing a lexical entry and/or its lexically encoded features. An N400 effect (i.e., a significant difference in the N400 amplitude) observed in response to two distinct lexical items is then a reflection of the relative difficulty in accessing their respective lexical entries and the features encoded therein. If we assume that, of the features that are encoded in the lexicon, some of these are semantic in the sense that they mediate, somehow, between syntax and thought (i.e., between a word and a concept), then the N400 can be taken to reflect contextual influences on the process whereby a hearer goes from word recognition to concept retrieval—and this would be inconsistent with the encapsulation hypothesis under discussion.

So stated, there are of course many ways in which the encapsulation proponent might respond to this challenge. First, one could argue against the lexical view of the N400 effect. If the N400 response is a reflection of post-access processes of integration, then it reflects processes that are functionally downstream from the kind of word-to-meaning transformation that is under discussion here. And, as indicated above, there are some reasons to prefer the integration view of the N400 over the lexical view. As noted, one of the most

compelling arguments offered for the integration view arises from early semantic priming studies, in which no N400 effect was observed after the target (e.g., 'tea') when the priming word ('coffee'/'chair') was masked below the level of conscious awareness. However, as mentioned above, later masked priming studies have found an N400 effect, provided that the interval between priming word and target was short enough.

Still, there remain at least two good reasons to prefer the integration view. First, the N400 effect has often been said to occur too late to reflect processes of lexical access (Sereno et al. 1998; Hauk & Pulvermuller 2004; Hauk et al. 2006). Second, and much more compellingly, N400 responses are observed for many non-linguistic stimuli. An N400 response can be detected following most stimuli that are, very broadly speaking, potentially meaningful—so, not only words, but also non-words that are pronounceable (i.e., pseudo-words like 'blicket;' Bentin et al. 1985; Rugg & Nagy 1987), faces (Barrett et al. 1988; Barrett & Rugg 1989), and pictures (Barrett & Rugg 1990; Holcomb & McPherson 1994; for a summary, see Lau et al. 2008: 920). Given this observation, the integration view has an intuitive advantage over the lexical view of the N400. Since our minds do not keep lexicons for pseudo-words or for pictures, it is hard to see how the lexical view of the N400 would explain these findings.

The lexical view seems most attractive when we focus on N400 responses to actual linguistic stimuli—i.e., it is plausible to think that the increased N400 in response to 'socks' as compared to 'cream' reflects differential processing costs involved in retrieving a specific entity stored in long term memory. However,

when we turn our attention to N400 responses to non-linguistic stimuli, the lexical view begins to appear less plausible. Indeed, given that the faces and pictures presented to experimental participants in the non-linguistic EEG studies are novel, not familiar, and so not stored in long term memory, it is hard to see how any kind of retrieval-based account of the non-linguistic N400 responses could be correct. The metaphor of retrieval or access seems inapt when it comes to the non-linguistic cases.

Alternatively, the proponent of Fodor-like encapsulation for semantics could accept the lexical view but insist that the N400 is a reflection only of the process of lexical access itself—i.e., word recognition—and not a reflection of whatever processes result in activating the features that are stored in the lexicon along with the word itself. In this way, one might accept the lexical view of the N400, but then seek to explain the N400 data in a way that preserves the thesis that the relevant semantic features are retrieved automatically, given word recognition. On this line of thought, the contextual influences that are reflected in the N400 effect are influences on word recognition alone, with activation of the word's lexically encoded semantic features coming along "for free." While lexical access is a contextually modulated process and hence not encapsulated, the activation of the features lexically encoded with the accessed lexical item is not contextually modulated, and so encapsulated in the relevant sense. This cedes some ground (i.e., the process of word recognition) to proponents of unencapsulated processing; however, it preserves the encapsulation hypothesis for content resolution, or concept retrieval, given recognition of a word. Of course, you might

accept the lexical view but somehow reconcile it with the N400 data, without accepting that lexical access itself is unencapsulated. However, as noted above (chapter 1), a meta-analysis of the relevant data supports the conclusion that lexical access is not encapsulated in the strict, Fodorian sense (Lucas 1999).

So, there are a few ways in which the encapsulation hypothesis might be defended in light of the experimental observations reviewed here. The proponent of encapsulated semantics can concede that, in many ways, practicing psycholinguists are correct: we do seem to observe top-down influences of extra-linguistic context on nearly every stage of utterance interpretation, to potentially include even word recognition. However, none of the relevant observations show contextual effects on the process by which, having recognized a word, a hearer activates its lexically encoded semantic properties. After all, such a narrowly semantic process of word-to-meaning transformation might occur alongside contextually modulated processes of spreading conceptual activation, such that even mind-bogglingly early activation of (e.g.) the concept CUSTOMER (and its activation being significantly greater than that of SANDWICH) in response to the linguistic input 'ham sandwich' simply could not count as evidence against the encapsulation hypothesis. Nor could the fact that one quickly adapts to understanding uses of 'peanut' as involving reference to an animate entity show that the operations of the narrowly semantic processor over 'peanut' as input are subject to contextual influences.

We see again that the encapsulation hypothesis under discussion here has a very narrow scope, and that its discussion is therefore quite delicate. It concerns

what I have called the "smallest" semantic process—that process which, given recognition of a word, retrieves something that is composable and that might mediate, somehow, between syntax and thought. As seen here, when we heed Fodor's central piece of advice concerning putative evidence against encapsulation—in short, by minding the locus of the effects in question—we can preserve the encapsulation hypothesis for this very small semantic process, even in light of an experimental record that shows quite a lot of top-down influence on other aspects of utterance interpretation. Indeed, it becomes hard to see what kind of evidence could, even in principle, undermine the encapsulation hypothesis for content resolution. This might lead one to worry that, at the level of meaning retrieval given a word as input, the encapsulation hypothesis is simply too fragile to permit of testing. And this could be either a limitation in practice, given the methods and results currently on the table, or a limitation in principle. If the latter, one might then suspect that the encapsulation hypothesis for word meaning retrieval is unfalsifiable—a worry I discuss below in chapters 4 and 5. For now, however, there is at least one other potential source of evidence that might be pertinent to the encapsulation question: data from neuroimaging studies.

3.4 Modularity and the neuroanatomy of semantics

When it comes to the relevance of the N400 data to the encapsulation question, part of the problem is that timing data is in general not well-suited to testing modularity hypotheses. For one thing, an encapsulation hypothesis for a given process does not entail that no unencapsulated processes occur in parallel. Fodor

suggests that encapsulated processes are much faster in their moment-by-moment operation than unencapsulated processes—"mind-bogglingly" so (Fodor 1983: 61). This claim about relative timing seems plausible when we compare, for example, edge detection with decision making. However, many component processes of utterance interpretation that are clearly inferential (and hence unencapsulated) are quite fast. As mentioned briefly above (section 2.5), visual world eye tracking studies indicate that hearers and readers resolve pronominal reference within about 200ms of stimulus onset (Arnold et al. 2012).[27] Since unencapsulated processes of pragmatic inference can occur in parallel with the semantic processes that are hypothesized to be encapsulated, and since these parallel pragmatic processes can themselves be mind-bogglingly fast, it is quite difficult to develop arguments, on the basis of timing data alone, either for or against the encapsulation hypothesis.

Accordingly, it would be helpful if we could supplement timing data with neuroanatomical data in order to ascertain whether the smallest semantic component of utterance interpretation (i.e., retrieval of a composable mental entity given word recognition) is encapsulated or not. Cognitive processes are realized by underlying neuroanatomical structures. The encapsulation hypothesis, following Fodor, is paired with the additional hypothesis that encapsulated modular operations are supported by dedicated (though possibly spatially distributed) neural systems (Fodor 1983: 98-99). Accordingly, one promising way

[27] See Green et al. 2020 for a related study establishing a similar conclusion for implicit reference, as in 'Al talked to Beth before ___ leaving,' resolution of which is shown to be equally rapid as compared to explicit reference.

to test the encapsulation hypothesis for a given cognitive function is by first ascertaining which brain regions support said function, and then experimentally testing whether or not the activity in said region is modulated as a result of top-down, extra-modular pressures of context and general cognition.

In fact, this is the strategy pursued in some of the more compelling arguments that have been offered against encapsulation of sensory processes. As noted in chapter 2, there is general agreement about the neural basis for early visual processing, i.e., construction of a geometrical representation of a scene (see Pylyshyn 1999, Ogilvie & Carruthers 2016: 722). Signals from the retina are first projected to a region in the back of the brain, known as V1, and then fed forward to other early visual subsystems that process the signals for information about contour, color, motion (V2-5). As Ogilvie and Carruthers state, "we know enough … to know that this network of regions serves to realize the early visual system. Evidence that the representational properties of these regions are modulated top–down by external cognitive factors will thus be evidence that early visual processing is not encapsulated" (2016: 724).

So, having identified the brain regions that support early visual processing, Ogilvie and Carruthers then point to several experimental observations that suggest top-down influences of external cognitive factors on the activity in these regions. For example, visual imagination involves top-down influences of goals on mental imagery and is supported by the same brain regions that support exogenously caused visual processing (see, e.g., Kosslyn 1994). Since top-down modulation of the regions recruited in early visual processing is possible for

endogenously caused perception, it is not immediately clear why it would be impossible for processing of bottom-up input, in cases of exogenously caused perception, to be likewise susceptible to top-down influence (Ogilvie & Carruthers 2016: 728). Ogilvie and Carruthers appeal to several other experimental observations to bolster their case. Task-related effects on activity in the early visual areas have been observed, suggesting again that goals can have a top-down influence on the brain regions that subserve processing of bottom-up visual input (Ogilvie & Carruthers 2016: 730-733). It has also been shown that expectations modulate activity in early visual processing areas and do so independently of attentional effects (2016: 733-736).

Ogilvie and Carruthers of course concede that one might defend encapsulation, at least in principle, by raising objections to the arguments just briefly described; most of these objections appeal to one of Fodor's ground rules. Ogilvie and Carruthers respond to each of these objections, concluding that the case against encapsulation of vision is stronger than the case in favor (2016: 739-740). I will not review the details of this discussion here. The purpose of this quick review of the anatomical case against encapsulation of vision is analogical: the encapsulation hypothesis for vision has been attacked in many ways, most of which are unsuccessful for the very reasons discussed throughout this chapter. The neuroanatomical strategy deployed by Ogilvie and Carruthers in their attack on encapsulation of vision is uniquely promising, in that it avoids the pitfalls that thwart other lines of attack. So, my goal here has not been to defend a particular conclusion about the encapsulation of vision, but to highlight what I take to be

one of the more promising strategies for falsifying it, so that we might consider how this strategy could be deployed in discussions of semantics and pragmatics.

The strategy being considered here is to first identify the brain area(s) that are involved in "early" semantic processing (i.e., retrieval, given recognition of a word, of some composable mental entity), and then to assess the experimental record to see if there is evidence of top-down modulation of activity in that area (or those areas). Several neuroanatomical models of semantic processing have been proposed. The models I review here are based on careful analysis of observations from a range of experimental measures. Still, this strategy will not lead to strong conclusions, in part because we know less about the neuroanatomy of sentence processing than we do about the neuroanatomy of (e.g.) visual processing. The models of semantic processing that we do have are still highly speculative, and widespread consensus has not been achieved. Additionally, extant models of sentence processing rarely include any specific component that corresponds to the kind of narrow semantic processing that is ultimately the focus here—the smallest semantic process whereby some kind of composable mental entity is retrieved given word recognition. Again, though, my rather modest goal here is to show how a particularly promising strategy for attacking encapsulation (first identify the supporting neural anatomy and then test whether its activity is susceptible to top-down manipulation) might be deployed with regard to semantic processing, and to take at least a few initial steps in deploying that strategy.

Lau and colleagues (2008) propose a model of semantic processing that integrates findings from several different experimental methods, including EEG,

magnetencephalography (MEG), fMRI, and studies of patients with brain lesions. On their model, the posterior middle temporal cortex (PTC) "mediates long term storage of and access to information associated with lexical representations;" the anterior temporal cortex (ATC) and angular gyrus (AG) "support integration of lexical input into larger units under construction, such as the syntactic structure, the propositional semantics and the discourse model;" and, finally, the inferior frontal cortex mediates "controlled selection and retrieval of lexical representations" (Lau et al. 2008: 923).

Where should we locate the smallest semantic process—i.e., retrieval, given word recognition, of some composable mental entity—within this model? Regarding the PTC, Lau and colleagues write:

> Which aspects of lexical representations are stored here is unknown. One possibility is that this stores not semantic information, but rather lexical representations that interface with a semantic network that is distributed across brain regions…However, it is also possible that this area stores the conceptual features that are associated with lexical representations. (Lau et al. 2008: 922)

This last suggestion, that the PTC might be the brain area that "stores the conceptual features that are associated with lexical representations" can be interpreted either in a narrowly linguistic way (such that the conceptual features in question are not merely associated with the words but are in some sense features of the words, or uniquely determined thereby), or in a broader way. Interpreted in the narrow sense, we could take this model to suggest that the PTC serves to store the information that a word encodes a particular concept[28] (or perhaps a list of

28 As noted above and implicitly acknowledged in the parenthetical, this is one (rather simple) way of thinking about the relation between words and concepts, to be revisited below in chapter 4.

concepts, or a pointer to a region in conceptual space). Interpreted in the broader sense, the model would have the PTC serving as storage space for non-linguistic, merely associational relations among words and concepts.

The anterior inferior frontal cortex is then involved in "retrieval of lexical information," and is therefore another candidate (along with the PTC) for the anatomical basis of the smallest semantic process, i.e., retrieval of a composable mental entity that might mediate between linguistic input and functionally downstream stages of utterance interpretation. The model of Lau et al. 2008 does allow for functional interaction between these areas, such that the narrowly linguistic information retrieved by the IFC given recognition of a word might then provide input to the operations of the PTC, to include activation of concepts associated with the word.

This, again, is just one possible way of reading one proposed model of semantic processing. It is also a speculative reading, since the model under consideration here does not make explicit where the smallest semantic process is located. Another model of semantic processing is defended in Matchin & Hickok 2020. On this model, the posterior middle temporal gyrus (pMTG) interconnects systems for speech perception in the temporal parietal lobe with "conceptual-semantic systems" in the inferior parietal lobe, where the latter regions are specialized for semantic comprehension in particular (Matchin & Hickok 2020: 1482). The conceptual-semantic systems in question include one that stores

---

For now, this simple idea (that a word encodes a concept) suffices for a discussion of the question as to whether the smallest semantic process is encapsulated, since this is largely orthogonal to the question about the nature of its outputs (though see chapter 4 for more discussion of the latter question).

knowledge of entities in the anterior temporal lobe (ATL)[29] and one that stores knowledge of events in the angular gyrus (AG) (Matchin & Hickok 2020: 6, citing Binder & Desai 2011).[30]

Here, again, we have a model of semantic processing that includes no specific mention of the smallest subcomponent of semantic comprehension: retrieving some composable mental entity given recognition of a word. The ATL and AG are hypothesized to store mental entities of a certain kind—namely, information about things and events (respectively). Such mental entities are, at least in principle, candidates for the output of the smallest semantic process. However, we have good reasons, independently of neuroanatomical considerations, to deny that all the information I possess with regard to dogs is the output of a narrowly linguistic process over 'dog' as input. As mentioned already,

29 The ATL was initially hypothesized to provide the neural basis for syntactic processing, since this region showed increased activation for sentences versus word lists, music, and sequences of environmental noises (e.g., Mazoyer et al. 1993). However, subsequent studies did not replicate these early findings, instead pointing to the IFG and pMTG as the basis for syntactic processing (e.g., Pallier et al. 2011). Later imaging studies then found that the ATL responds to "conceptual-semantic or discourse properties regardless of syntax" (e.g., Fletcher et al. 1995), but does not "respond to manipulations of syntax independently of semantics" (e.g., Rogalsky & Hickok 2008). Adding to the case, patients with ATL damage show considerable conceptual-semantic deficits, while their basic sentence comprehension (when lexical demands are minimized) and production capacities are unimpaired (e.g., Hodges et al. 1992), with the same sensitivity to grammatical violations as healthy individuals (e.g., Grossman et al. 2005). See Wilson et al. 2014 for a detailed overview of the relevant data (Matchin & Hickok 2020: 1482-1483).

30 The object/event distinction is often argued to be of fundamental psychological importance, in part based on experiments that show that children and adults tend to group objects in one of two ways: based either on shared individual attributes (such that DOG and CAT are grouped together) or based on how they interact with other objects (such that DOG and BONE are grouped together) (see, e.g., Lin & Murphy 2001). The localization of this distinction in the ATL versus AG is based largely on, first, the observation that a greater tendency for taxonomical errors (using 'dog' instead of 'cat') is reported in patients with stroke-induced ATL damage, while thematic errors ('dog' for 'bone') were associated with AG damage (Thompson et al. 2007); and, second, the observation that stimuli involving thematic relations result in increased AG activation compared to stimuli with taxonomic relations or non-thematic feature combination (e.g., Kaleinine at al. 2009); by contrast, we see increased activation in the ATL, not the AG, when the stimuli involve enriched taxonomic representations of an entity, such as in adjective+noun combinations like 'red boat' (see Pylkkänen 2015 for general overview) (Matchin & Hickock 2020: 1484-1485).

not every thought or concept that is occasioned by an utterance of a word is a direct causal result of recognition of the word by itself, independent of any extra-linguistic contribution. The kind of narrowly linguistic process under consideration here is one that could, however, mediate between word recognition and activation of the extra-linguistic information stored in the ATL or AG.

Until models of semantic processing are developed that posit a specific brain area as the neural basis for the smallest semantic process, the strategy I have borrowed from discussions of visual encapsulation (i.e., from Ogilvie & Carruthers 2016 and the authors they cite) cannot be pursued much further with regard to semantic encapsulation. However, I have sketched here how one might begin to apply, with regard to semantic processing, a strategy for attacking encapsulation that has proved to be promising in other, somewhat better understood domains. Future work in psycholinguistics might provide the necessary evidentiary basis for the kind of fine-grained models that would be needed in order to pursue this dialectical strategy further.

3.5 Interim conclusions

So, it seems that, in principle, behaviorally rich measures (such as explicit verbal responses to questions) and temporally rich measures (such as EEG) are ill-suited in principle for testing encapsulation hypotheses. Behavioral measures generally leave too much time in between stimulus onset and response generation, such that either post-modular or parallel but extra-modular processes might contribute to the response. And, given the possibility of rapid parallel processing by extra-

modular systems, timing data is unlikely to settle questions about encapsulation. Although the neuroanatomical strategy considered above is promising, in that it could at least in principle afford an answer to the encapsulation question, in practice the extant models of sentence processing are generally not fine-grained enough, in that they do not include the smallest semantic process as an explicitly listed component. We might therefore seek alternative sources of evidence that might help us to decide the encapsulation question, as I do in the next chapter.

Chapter 4: Polysemy and encapsulation

Another promising strategy for assessing the encapsulation hypothesis for semantic processing would be to pay close attention to the phenomenon of polysemy. Consideration of polysemy is useful when we ask what the semantic processor might yield as output given a disambiguated word form as input, since it allows dissociation of the functions of word recognition and concept retrieval. Given as input '$bank_1$,' what might an encapsulated semantic processor yield as output? If we knew the answers to this question, we could see whether one of them is best supported by the empirical record. If so, we might take this to provide some support for the encapsulation of a semantic process that mediates between word recognition and concept retrieval.

As argued in what follows, the Fodor-like encapsulation of the semantic processor would rule out certain options—e.g., that it might sometimes yield INSTITUTION and sometimes LOCATION, depending on extra-linguistic context. A few different options are available regarding the outputs of an encapsulated semantic processor given polysemous input. To simplify things considerably for now, suppose we thought the best candidate for the output of a narrowly linguistic process over a word like '$bank_1$' was a chimerical combination of concepts, a list-like entity that might include INSTITUTION and LOCATION as parts. Then, if we found that both of these concepts were activated given recognition of '$bank_1$,' even in contexts where one sense is much better supported by extra-linguistic context, this might be taken to provide some support for the encapsulation

hypothesis. Before pursuing this strategy on behalf of encapsulation, however, we must sort through the candidate entities for the output, given a polysemous word as input, of the smallest semantic process.

In many ways, this discussion follows recent work by other cognitivist philosophers of language (Vicente 2018; Carston 2019, 2020). However, I point to some crucial shortcomings in the dialectic thus far. Agustin Vicente (2018) argues in effect that encapsulation requires that some polysemes, such as '$bank_1$' (a 'regular' polyseme; see section 4.2 below), must be mapped onto a chimerical mental entity that combines the concept of a physical location with the concept of a social institution (along with a few other concepts). However, Vicente's argument begs the question against the view that '$bank_1$' makes available not a discrete list of concepts, but rather something like a pointer to a vague region in conceptual space (Carston 2012; cf. Ruhl 1989; Pietroski 2018).

As I argue, the extant observations and considerations do not adjudicate between chimeras and pointers. In short, then, despite what has been suggested in the literature (Harris 2020), the encapsulation hypothesis does not entail (even in conjunction with considerations about polysemy) any one particular view about the outputs of the smallest semantic process. And, also contrary to certain suggestions in the literature (Vicente 2018), none of the available data or arguments adjudicate among the two main candidate theories regarding what those outputs might be, chimeras or pointers. So, what is needed at this stage is careful reflection about what testable predictions could differentiate the two

theories. I conclude this chapter by attempting to articulate some differentiating predictions along these lines.

4.1 What encapsulation does not require

It is sometimes argued (or, in some cases, merely suggested) that a particular approach to word meaning is best suited to the encapsulation hypothesis. For example, it has been suggested that the encapsulation of a semantic system requires that its outputs be highly general concepts (Borg 2004). But it has also been argued that encapsulation requires instead that words be mapped to non-conceptual entities, such as constraints (viz., on expressible concepts; see, e.g., Harris 2020).

These two arguments include more or less the same premises but arrive at mutually exclusive conclusions. As I will show here, both arguments are unsuccessful, since the encapsulation of semantics is compatible with several different kinds of output. A word might context-invariantly afford something conceptual and specific, or conceptual and under-specific, or perhaps something entirely non-conceptual. None of these options is ruled out by encapsulation. From the encapsulation hypothesis alone, we can infer very little about the output of semantic processing of a polysemous word.

4.1.1 The argument from encapsulation to minimalism

According to a version of semantic minimalism, given a certain word as input, the semantic system context-invariantly outputs a highly general concept that is

uniquely determined by the word in question. Given a sentence as input, the semantic system context-invariantly outputs a certain highly general but truth-evaluable thought. There is a line of argument from the encapsulation hypothesis for semantics to this kind of minimalism, but it is not a very good argument. After reviewing this argument, I will show why it is unsuccessful. However, while the argument for minimalism is unsuccessful, there is a related, but more modest, conclusion to be drawn from consideration of polysemy and encapsulation.

Having argued in favor of the encapsulation hypothesis for semantics in her 2004, Emma Borg goes on to argue that a version of minimalism is particularly well-suited to this hypothesis. One of Borg's explicitly stated goals in her 2004 defense of minimalism is to show that it allows linguistic meaning to be "a proper part of a modular language faculty" (2004: 74), and this is because minimalism is a formal theory of meaning, in the sense that (i) its operations are computational and syntactically driven, and (ii) semantically encoded context-sensitivity is narrowly restricted to a limited range of objective contextual features (2004: 32-33). As Borg argues, "a formal semantic theory alone can be part of a Fodorian modular language faculty" (2004: 75). As a formal semantic theory, then, minimalism is well-suited as an account of the semantic component of the language faculty.

However, this argument from encapsulation to minimalism is unsuccessful. As Borg acknowledges, the very same considerations she offers in favor of minimalism could also be taken to support a Relevance Theoretic account on which the language faculty contains not a complete formal semantic theory, as the

minimalist requires, but rather a fragment thereof (Borg 2004: 108). For some Relevance Theorists, the semantic decoding component of utterance interpretation is a fragment of a formal semantic theory: given a sentence as input, it yields not a truth-evaluable thought, but an incomplete logical form, with derivation of truth-evaluable content being a post-linguistic process (Sperber & Wilson 1986; see discussion in chapter 1). That is, a conceptual but non-truth-conditional formal semantic theory is perfectly compatible with the modularity hypothesis for semantics.

While proponents of the constraint-based approach to semantics have not always bothered to do so, it can be worked into a formal compositional semantics, matching the more traditional versions of compositional semantics in terms of descriptive adequacy and explanatory power. Over an extended body of work, Paul Pietroski has shown that a formal semantic theory that trades not in truth-conditions, nor even in concepts, but trades instead in instructions to retrieve and combine concepts, can rival more traditional semantic theories in terms of descriptive adequacy and explanatory power (see, most recently, Pietroski 2018).[31] As discussed briefly in chapter 1 and again in more detail in what follows, such non-conceptual accounts of word meaning are often explicitly motivated by appeal to polysemy. There is no single concept with which '$bank_1$' stands in a one-to-one relationship. So, if the relation between a word and its semantic output is one-to-one, then the word must be paired semantically with something non-conceptual.

[31] Harris (2020) has also recently defended a constraint-based but nonetheless formal semantic theory within a cognitivist framework.

Hence, the considerations Borg points to regarding encapsulation fail to rule out a conceptual but non-truth-evaluable semantic theory, of the sort defended by early Relevance Theorists (Sperber & Wilson 1986). The line of argument here also neglects the kind of non-conceptual approach defended in later work by Relevance Theorists (e.g., Carston 2012), and by cognitivists who are not Relevance Theorists as well (e.g., Harris 2020).[32] That the processor is encapsulated does not preclude the possibility that, given a word as input, the processor yields as output a non-conceptual mental entity, such as a constraint on expressible concepts, or an instruction to fetch a concept at a particular mental address.

Finally, a chimeric semantics is compatible with the encapsulation hypothesis (see, e.g., Langacker 1991; Asher 2011; Vicente 2018; Carston 2020; the view has antecedents in the work of Pustejovsky; see Pustejovsky 1995). This too is a kind of non-conceptual semantics, in that a word is paired not with a concept but with, in effect, a list of concepts. As with pointer semantics, chimeric semantics is generally motivated by consideration of polysemy, especially so-called regular polysemy, as exhibited by '$bank_1$.' On the chimeric view, rather than encoding a concept, or a pointer to a region in conceptual space, a word like '$bank_1$' encodes a structured object consisting of the various concepts that it can

32 I do not include Pietroski here because his view is strictly speaking about competence, not performance. So, for example, Pietroski makes explicit no commitment to the hypothesis that a processing system context-invariantly translates a word-form to an instruction to fetch a concept at a mental address. And, as a competence theory, Pietroski's instruction-based semantics does not entail any such commitment. By contrast, the encapsulation hypotheses defended by Borg, Harris, and the Relevance Theorists are explicitly said to concern a processing module. (See chapter 2 above and chapter 5 below for some relevant discussion of the distinction between competence and performance in this domain.)

be used to express. Among the concepts on the list, some are mutually exclusive (e.g., no one thing is both a physical location and a social institution). Hence, many open-class words are paired, on this view, with combinations of concepts of mutually exclusive ontological categories—or what I call 'chimeras.'

On the chimeric account, across all contexts of use, '$bank_1$' is mapped by the semantic system to the same chimera. On the pointer theory, '$bank_1$' is context-invariantly mapped to a pointer to a region in conceptual space, a region that might have vague boundaries. As discussed below in more detail, the key difference here is that lists (or chimeras) are necessarily discrete things, whereas a region in conceptual space might be vague; i.e., a region might lack a sharp boundary. This allows, on the pointer view, the possibility that one might (in one way of conceptualizing the pointer-based hypothesis) follow the instruction afforded by the word and activate a concept in a particular region, but a concept that is on the region's penumbral edges. No such correlate of this vague periphery exists on the chimeric theory: lists of concepts do not have peripheries.

Much of this is very metaphorical. For now, the lesson is that various non-conceptual approaches to word meaning are perfectly compatible with the Fodor-modularity of semantic processing. So, there is no compelling argument from encapsulation to a minimalist account of word meaning. That said, encapsulation does allow us to draw one conclusion that minimalists will enthusiastically endorse: the output of an encapsulated semantic processor, given a word as input, is generally *not* going to be identical with any component of a veridical representation of what a speaker means to express in using that word. In other

words, encapsulated semantic processing cannot eventuate in attributions of speaker meaning, even in the direct sense. What a speaker means is simply too context-sensitive to be the result of encapsulated processing; it is either the purview of non-modular processing (on a Fodorian account), or modular processing that is minimally encapsulated (on a massively modular account). Hence, encapsulation does, it seems, support minimalism over other accounts on which a word is mapped to a concept, as consideration of polysemy makes especially clear. If the smallest semantic process maps a given word onto some uniquely determined concept, consideration of polysemy indicates that said concept must be highly general with respect to the many specific senses that a speaker can use the word in question to express.

4.1.2 THE ARGUMENT FROM ENCAPSULATION TO CONSTRAINTS

Dan Harris offers the following argument from the encapsulation hypothesis to the conclusion that the outputs of semantic processing must be constraints. First, as already discussed, semantic processing is taken by Harris to be the work of a modular system that is encapsulated from general cognition. Second, it is argued "that content resolution is a central-cognitive process" (2020: 14). According to Harris, given these two claims, it follows that "semantics can only deliver constraints, not contents" (o.c.). In other words, because semantic processing is the work of a modular system encapsulated from the workings of central cognition, and because mindreading is a central cognitive process, the result of semantic processing cannot be influenced by mindreading. Since resolving

propositional contents requires some amount of mindreading and general pragmatic inferential work, representations of propositional contents cannot be the output of semantic processing. Instead, semantic processing must output something like a constraint over expressible contents.

Again, here I am granting that semantic processing is the work of an encapsulated module (though, as argued in chapters 2 and 3, the case in favor of this hypothesis is not particularly strong when it comes to the smallest semantic process: retrieving a composable mental entity given word recognition). And I agree with Harris' second premise, "that content resolution is a central-cognitive process" (2020: 14), mostly for the reasons that Harris himself notes, citing many others—but with an important qualification. The second premise is uncontroversial, and not question-begging, only given one possible understanding of the notion of content resolution. On a different way of understanding 'content resolution,' the second premise is question-begging and arguably false. The distinction I rely on here is a very familiar one: that between the content of a sentence *S* and the content of a speech act performed in uttering *S* (see, e.g., Grice 1968; Kripke 1977; Neale 1990).

On one interpretation of Harris' argument, the second premise concerns the content of a speech act: the content of a speech act performed by uttering a sentence *S* is resolved not by the work of a semantic system that is encapsulated from general cognition, but rather by way of central cognitive processes such as those involved in mental state attribution. Of course, attribution of speech act content may in part involve the output of semantic processing of the linguistic

input; but the output of semantic processing is not itself an attribution of direct speech act content. Working out what someone has asserted requires a hearer to rely not only on her semantic competence, but also on a range of other cognitive capacities, many of which are not the work of Fodor-modular subsystems (cf. Allott, forthcoming). Direct speech act content attribution often requires inferring the reference of indexicals, pronouns, and other obviously context-sensitive expressions; such inferences may in principle draw on an unlimited range of information. So, on this interpretation, the second premise of Harris' argument is uncontroversial: speech act content resolution must be the result of central cognitive processes (or, for massive modularists, the result of processes that are modular in a non-Fodorian way, involving only a limited form of encapsulation). This is precisely the kind of point that minimalists often make: we must not confuse aspects of speaker meaning attribution with aspects of semantic understanding, since the latter is strictly a linguistic matter, and the former is much more than that. And, as noted above, this allows us to conclude that, indeed, if semantic processing is Fodor-modular, then it does not eventuate in speaker meaning attributions.

Harris' argument, again, consists of two premises: semantic processing is the work of a modular system that is encapsulated from general cognition and "content resolution is a central-cognitive process" (Harris 2020: 14). However, an uncontroversial conclusion follows from this pair of premises, insofar as we understand 'content resolution' to refer to the attribution of speaker meanings. That narrowly semantic understanding does not by itself suffice for an

understanding of what a speaker has asserted on a given occasion is, at this point, a platitude. Again, consider pronouns and the like, but also polysemy, for example.

However, Harris states that this uncontroversial claim is not the one that he wishes to defend. In responding to a potential objection, he states that "the kind of understanding that is relevant to my argument [...] is successful semantic composition," and not "understanding a speaker" (2020: 11). Again, the issue of semantic composition is a red herring here: since it is generally agreed that mode of semantic composition does not vary as a function of extra-linguistic context, the encapsulation of semantic composition comes easy. What is subject to debate is whether something like meaning retrieval might be the work of an encapsulated processor. But if we take Harris' argument to support a conclusion about expression contents, we must read the second premise as likewise a claim about expression contents, not speech act contents, on pain of equivocation. And, so understood, the second premise begs the question against the minimalist (and perhaps also the chimeric semanticist). Semantic minimalism is clearly a version of the content-based approach to semantics: minimalism is about as austere and uncompromising a version of content-based semantics as one can find. And the core claim of minimalism is compatible with the claim that semantic processing is modular and encapsulated from central cognition. That semantic processing is the work of an encapsulated system does not preclude the possibility that the system outputs minimal concepts. Borg and others have shown the compatibility (in principle) of minimalism and Fodor-modularity, as noted above.

Harris' argument either supports an uncontroversial conclusion (that speech act content attributions are not the result of the operations of an encapsulated semantic module), relies on a question-begging premise (that expression content resolution could not be the result of the operations of an encapsulated semantic module), or equivocates in its use of 'content.' Minimalists can allow that resolution of expression contents (grasping the thought that is the minimal content of the sentence uttered, or grasping the minimal concept that is the content of a word) is the work of an encapsulated semantic module, and also accept that attribution of speech act content is the work of central cognition, but deny (as they characteristically do) that speech act contents are expression contents, and thereby reject the conclusion that semantics cannot output contents.

4.1.3 WHAT ENCAPSULATION REQUIRES, GIVEN POLYSEMY

I have argued here that, even if we grant the encapsulation hypothesis for semantic operations, to include not only composition but also retrieval of the units for composition, this tells us very little about the nature of those units. Several different accounts of the outputs of semantic processing, given a word as input, are compatible with the encapsulation of said processing. All that is required by the Fodor-style encapsulation hypothesis under consideration is that semantic processing of a word as input is not susceptible to top-down, extra-modular influences. The output of such an encapsulated process could (at least in principle) be a highly general concept, as the minimalist maintains; or it might be a list of different concepts, as the chimeric semanticist proposes; or it could be

something vaguer, like a pointer to a region in conceptual space, an instruction to fetch a concept, or a constraint on possible truth-conditional contributions or expressible concepts. Each of these options is compatible with the encapsulation hypothesis.

Since the encapsulation of semantics leaves open a disjunction of possibilities regarding its outputs, we must look elsewhere for evidence that might rule out some of these options. Consideration of polysemy is useful here. Polysemy, again, is the phenomenon whereby a single word can be used to express any member of a set of related senses or concepts. As argued in the preceding chapter, because polysemy is a property of a single word, it is not easily explained away, in Fodorian fashion, as a pre-semantic or otherwise extra-modular phenomenon. An experimental literature on the processing of polysemous words is available to guide our choice concerning the outputs of an encapsulated processor. The authors of these studies generally assume the encapsulation hypothesis, rather than testing it. Such studies are therefore not very useful when it comes to assessing the encapsulation hypothesis itself. However, the experimental literature on polysemy, precisely because it generally takes for granted the encapsulation hypothesis regarding the semantic processing system, offers a promising strategy for determining what such a system might yield as output.

This strategy has been deployed before, but with mixed results. For example, Agustin Vicente (2018) offers an empirically guided sequence of arguments regarding what sort of mental entity a polysemous word might afford

across contexts of its use. Carston 2020 includes a similar discussion, arriving at similar conclusions. With this much, I am in agreement with Vicente and Carston: the available experimental record suggests that the best account of polysemy overall is likely to be a pluralistic one that models the processing of different kinds of polysemy in different ways. Within a cognitivist and use-oriented framework, some examples traditionally classified as cases of polysemy might be reasonable candidates for assimilation to homophony. But this strategy is quite limited, and many cases of polysemy will resist such assimilation. Among these, different sub-categories may require different treatments (cf. Vicente 2018; Carston 2020).

In his 2018, Vicente additionally argues for the stronger conclusion that we *require* a chimeric semantics for regularly polysemous words like '$bank_1$' while acknowledging that, for irregular polysemes, a non-chimeric approach may be better suited. However, according to Vicente, the best account of regular polysemes has it that the semantic system maps such terms onto list-like, structured mental objects—chimerical combinations of concepts of different, often even ontologically mutually exclusive, categories. Carston (2020) argues for a similar conclusion, with many of the same premises, though with important differences in the overall line of argument. However, the arguments offered are unsuccessful. Before reviewing those arguments, however, a deeper discussion of the phenomenon of polysemy is in order, including the distinction between regular and irregular polysemy, as well as other taxonomic distinctions.

4.2 VARIETIES OF POLYSEMY: AN OVERVIEW

Polysemy is often defined as follows: a given word is polysemous just in case it has multiple distinct but related senses.[33] Polysemy is distinguished from homophony, in that the latter is a relation among distinct words that happen to share a pronunciation, whereas polysemy is a property of a single word. And polysemy is distinguished from generality, in that the relation between the senses of a polysemous word is not merely one of relative generality. So, for example, the relation between (or among) the senses of '$bank_1$' is different in kind than the relation between the senses of 'blue' and 'royal blue.'

The senses of polysemes bear various kinds of relation to one another, many of which generalize to patterns that are common both within and across languages. Below is a list (adapted from Srinivasan & Rabagliati 2015) of such regularized relations, together with examples in English, where the brackets give two possible continuations, each of which is more compatible, given general background knowledge about the world, with one of the senses of the polysemous subject expression.

Relation	Example
animal:meat	The chicken {walked quickly / tasted great}.
material:artifact	The glass {was strewn about the carpet / contained wine}.
vehicle:content	The book {is heavy/ is interesting}.
container:content	The pot {is metal / is boiling}.
producer:product	Dickens {was prolific / is boring}.

33 Of course, the word 'has' is itself polysemous (cf. Aristotle *Categories* 15b16-31). In what sense of 'have' does a polysemous word have its various senses? In the sense that I have my nose, but do not have the pen I hold in my hand—more or less. That will have to do for now, though I return to this question below in the concluding chapter 5.

building:institution	The bank {is made of concrete / is aggressively investing}.

This list of regularized patterns is not exhaustive. And a single word can have more than two senses, related to one another in more than just one of the aforementioned ways. Consider, once again, the go-to example, '$bank_1$,' which can be used to refer not only to physical buildings and the more abstract social institutions housed therein, but also to an institution's managers (as in 'The banks are coming by for lunch today, Mr. President').

Each of the examples of polysemy given in the above list is regular. A word with two senses related to one another in a certain way is a regular polyseme just in case the same language includes at least one other word with two senses that are related in the same way (Apresjan 1974: 16). So, for example, in English, we find the building:institution relation exhibited not only by '$bank_1$,' but also by 'school,' 'library,' 'White House,' etc. Regular polysemy is also quite robust across languages. All the relations listed above (along with several others) obtain among the senses of polysemous terms in at least fourteen other languages (Srinivasan & Rabagliati 2015).

Among regular polysemes, some have senses that are of mutually exclusive ontological categories; such regular polysemes are called 'inherent' (Pustejovsky 1995). For example, 'book' is inherently polysemous, as physical story vehicles and abstract story contents are entities of fundamentally distinct and mutually exclusive ontological categories. (If x is a story content, x cannot be a solid mass of mostly wood pulp.) Not all regular polysemes are inherent. Consider, e.g., the senses of 'chicken': individual organisms and their meat are both concrete,

differing in their countability, but not their fundamental ontological status. (If x is a countable chicken, then—alas for x—x can be a meal.)

Not all polysemous terms are regular in this way. Many are idiosyncratic and highly productive. Take, for example, the variety of senses with which 'line' can be used. We use one and the same word to refer to continuous lengths without breadth; to the most direct course between two locations; to queues of people or other things; to a length of rope or wire or chord; to certain strings of text; to performed utterances in films and plays; to marks on pages, sand, or skin; to certain sequences of thoughts, or events; and so on. The relations among these different senses of 'line' are not regular, in that they do not conform to a single, generalizable pattern, in contrast with the relations listed above. Instead, the relation obtaining among the senses of 'line' is merely one of family resemblance.

Though irregular (in the sense of Apresjan 1974), the senses of 'line' listed above are highly conventionalized, enough so that most of them get a mention in most English dictionaries. Of course, words can be used with senses that are not conventionalized in this way, but are instead "cooked up on the fly," or devised for a particular purpose on a particular occasion. There is no standing convention whereby one can use 'bulldozer' to express the concept of a forceful person. And yet in some contexts 'Jane is a bulldozer' can be used to directly assert that Jane is a forceful person (Wilson & Carston 2007). This kind of polysemy is ad hoc: the speaker intends to use a term to express a concept with which the word is not conventionally associated (often represented with an asterisk, *BULLDOZER), but in a way that the speaker expects the hearer to be able to grasp by way of

pragmatic inference. Appeals to ad hoc polysemy, however, are often hard to distinguish from appeals to metaphor, blurring what might be thought of as an important distinction (see further discussion in section 4.4.3, chapter 5).

In any case, there are various subcategories of polysemy. (To put the point in Aristotelian terms, 'polysemy' itself is said in many ways.) Polysemous words can be regular or irregular; and irregularly polysemous words can be conventionalized or ad hoc. Importantly, these distinctions may involve differences not in kind but in degree, and we should therefore expect to find examples that are tricky to categorize in one way or another (e.g., as conventionalized or ad hoc). Psychological investigations of the storage and processing of polysemous words in the communicating mind arguably ought to be informed by these distinctions. However, we should also be prepared to find that certain theoretically motivated divisions might not be neatly recapitulated in the communicating mind. Before pursuing this cognitivist discussion of polysemy further, it will be instructive to briefly turn off to the side and address the challenge that polysemy raises for the now orthodox, truth-conditional approach to semantics.

4.3 THE CHALLENGE FROM POLYSEMY FOR TRUTH-CONDITIONAL SEMANTICS

It has already been observed that regular, inherent polysemy raises a substantial challenge for the standard truth-conditional approach in formal semantics (Chomsky 1964; see Pietroski 2017 for an overview). In this section, I press this challenge. I then discuss a few responses. The goal here is primarily to argue that

the truth-conditional semanticist has available a limited range of viable options with regard to regular polysemy. As will be seen when we return to consideration of the cognitive processes underlying language use, roughly the same set of options is available, and the dialectic concerning the best account of polysemy is quite similar. So, the parallels should be informative.

On the standard approach in formal semantics, sentences are mapped to context-relative contents. Heim and Kratzer's 1998 textbook is the contemporary standard-bearer for this approach in formal semantics (with antecedents in Lewis 1970, Montague 1973, Kaplan 1989, Chierchia & McConnel-Ginet 1990). The content of a lexical item is something like an individual or a property, the latter being treated as a function from individuals to truth-values, relativized to various parameters, including a world and a narrowly defined set of contextual features. Given a rule of semantic composition, such as function application, we can derive from the contents of its constituent expressions the semantic content of a sentence. For example, the content of 'Bob speaks' is the proposition that Bob speaks, derived by function application of the content of 'speaks' to the content of 'Bob.'

In short, on the standard framework, semantics maps a sentence to a proposition, and lexical items to propositional ingredients.[34] That said, a sentence on its own does not determine a proposition, as is immediately clear given even brief consideration of sentences with pronouns, such as 'He speaks.' What propositional content this sentence expresses is a function of the intended referent

34 Alternatively, semantics maps a sentence to a truth-condition, and lexical items to their truth-conditional contributions.

of the pronoun. On the standard, content-based account, the semantic value of 'he' is an assignment-dependent referent. In a context *c*, some contextually determined mapping *g* of assignment-dependency encoding indices to referential contents determines the referent of 'he' in that context. 'He speaks' is then assigned as semantic value the proposition that is true at every world *w* at which the contextually determined referent of 'he' speaks. On this approach, semantics does not map sentences to propositions context-invariantly; sentences do however uniquely determine a proposition relative to a (narrowly defined) context.

However, consideration of polysemy suggests this cannot be quite right. A sentence containing a polysemous word does not uniquely determine a proposition, not even relative to an index of evaluation. Take (17), for example, where the numerical subscript indicates one resolution of the homophony of /bæŋk/.

(17) The bank_1 is collapsing.

On the now standard approach, given resolution of any homophony, the sentence (17), in addition to both a contextual specification of values for variables in its logical form and a mode of semantic composition, determines a propositional content. However, given the polysemy of 'bank_1,' this account appears to be inadequate. At least initially, it seems that we cannot say that (17) uniquely determines a proposition, not even when relativized to an index of evaluation.

This challenge is not addressed by acknowledging (correctly) that no semantic theory is expected to explain all possible variations in speaker meaning. It is at least plausible to think that a speaker of (17) might make a claim about a

particular building, or about a financial institution, or about such an institution's managers. Then again, it could very well be that there is no determinate matter of fact with regard to, not only whether, but also what, if anything, a speaker asserts on a particular occasion of use. Whether or not a speaker can assert this or that proposition, and if so when she asserts one rather than the other—these are not questions that a theory about sentence meaning needs to settle.

However, even acknowledging these qualifications about the explanatory scope of a formal semantic theory, the question remains: Which proposition does the sentence (17) uniquely determine relative to an index of evaluation? Is it a proposition including or about a physical building, or one about a financial institution, or one about a small group of people who run a financial institution? No answer is ready to hand in the truth-conditional semanticist's basic toolkit. One can insist that the truth-conditional contribution of '$bank_1$' is the same across (18-20):

(18) The $banks_1$ are giving out high interest loans, so watch out.

(19) The $banks_1$ in this town are all being condemned for demolition.

(20) The $banks_1$ are coming for lunch this afternoon, Mr. President.

But this seems like special pleading. After all, no single thing could give out a loan, be condemned for demolition, and have lunch. Generally speaking, people and institutions give out loans, buildings get condemned, and people (as well as other animals) have lunch. The conditions for (18-20) (assuming that sentences in

any sense have such things) therefore seem to differ, a difference that seems to be the result of the polysemy of 'banks$_1$.'

One might appeal to local co-text. In many cases, co-text does seemingly make plain the most probable sense of a polysemous word, as in (18-20) (or, even more locally, in 'savings bank' versus 'tall bank'). However, whether syntactic context makes certain resolutions of polysemy *more probable* than others sheds no light on the question as to whether a polyseme, plus its syntactic context, uniquely determines one of its senses. To make probable is not to uniquely determine, after all.

And, in many cases, the relevant contextual information is not made available by the local syntactic context, as with an utterance of (17).

(17) The bank$_1$ is collapsing.

(17) is one sentence. What proposition does it uniquely determine relative to an index? After all, (22) gives rise to a different judgment, contrasting with (21).

(21) The bank$_1$ is collapsing. The latest financial crisis is to blame.

(22) The bank$_1$ is collapsing. The latest monsoon is to blame.

Local, linguistic context does not always determine a particular resolution of polysemy. In fact, local co-text by itself never determines a particular resolution of polysemy. On an occasion of use, speakers intend, and hearers often manage to grasp, one or another of the senses of a polyseme. Local co-text can only guide the hearer's inference about the intended sense.

4.3.1 POLYSEMY IS NOT HOMOPHONY

Another solution would be to assimilate polysemy to homophony, since the latter is no threat to the standard truth-conditional approach in formal semantics. One might insist that what I have labeled 'bank$_1$' is in fact not a single word, but rather two distinct words, one that refers to institutions, the other to buildings. This kind approach is articulated in Katz 1972 (cf. Fodor 1998; Fodor & Lepore 2002). However, a traditional source of evidence for formal semantics—i.e., reflective judgments about the felicity of an utterance given a context—provides reason to reject such an approach.

Arguably, two conjoined predicates cannot be understood with respect to different resolutions of homophony, as suggested by (23):

(23) # After the storm, the banks had eroded much of their soil into the river and given out low interest loans to displaced homeowners.

The unacceptability of (23) shows that the two resolutions—RIVERBANK and FINANCIAL INSTITUTION—are senses of two distinct words. During utterance interpretation, sounds are mapped to lexical items. And this example shows that a single instance of the sound /bæŋk/ cannot be understood, in the interpretation of conjoined predicates, to express first the river- and then the finance-related sense.[35] This in turn suggests that those two senses are the senses of distinct lexical items.

35 An utterance of (23) could only work as a piece of wordplay. However, if this wordplay has the intended effect of amusement, it does so by exploiting the fact that the two senses are those of distinct words. A hearer is briefly led by the second predicate to consider the possibility of riverbanks giving out loans, then recalls that /bæŋk/ is the sound of an altogether different word, and (if the pun works) is amused by the jarring effect. As with many puns, the amusement results from coincidence—in this case, of the phonetic realization of distinct words with unrelated senses.

By contrast, the polysemy of '$bank_1$' can be exploited in sentences such as (24), where the conjoined predicates are readily understood with respect to two different resolutions. (The observation of the contrast between sentences like (23) and (24) is due to Zwicky & Saddock 1975, from which these examples are adapted.)

(24) The $bank_1$ has a weathered façade but a good reputation.

Unlike (23), (24) is judged to be acceptable. A single instance of the sound /bæŋk/ can be understood, in the interpretation of conjoined predicates, to express first the physical and then the institutional sense. This in turn suggests that those two senses are the senses of a single lexical item.

Anaphora provides a similar test (Cruse 2004). If a homophone is used with one sense, a subsequent anaphoric expression cannot be felicitously used with the alternate sense, as shown the unacceptability of (25).

(25) # This year, the banks overflowed during the monsoon season. They also invested heavily in offshore stocks.

By contrast, a polysemous term like '$bank_1$' can support use of an anaphor with the alternate sense.

(26) The bank has a weathered façade. But it still has a good reputation.

This observation reinforces the lesson of the co-predication test: the financial and riparian senses are paired with two distinct lexical items, while the institutional and physical senses are paired with one of these two items.

Distinguishing between polysemy and homophony is generally difficult. Standardly offered definitions are imperfect. For example, it is sometimes

suggested that the basic difference consists in the fact that homophony involves senses that are *unrelated* to one another, while the senses of polysemous terms are *related* (this is in effect the distinction given in your average dictionary). However, this is mistaken, since often enough the senses of two distinct homophones are in fact related to one another in a non-trivial way. For example, the financial sense of the word I have been labelling '$bank_1$' is in fact historically related to the fluvial sense of what I have been labelling '$bank_2$,' as both English words derive from a Proto-Germanic root meaning roughly 'shelf, bench' (Elbourne 2011). Hence, the existence of a non-trivial relation among the two senses is not sufficient grounds for diagnosing polysemy. Furthermore, a speaker nowadays might judge two senses to be related ('vinyl record,' 'written record'), despite the fact that (by the diagnostic tests listed above) they are the senses of two distinct homophones. Hence, relations among senses (whether intuitively "felt" or not) do not serve as a reliable guide when it comes to categorizing various instances as instances of polysemy or homophony.

However, in general, if a given indeterminacy can be exploited in the ways exhibited by (24) and (26), then it is not a case of homophony. Still, though, it does not follow that resistance to this kind of exploitation entails homophony. Irregular and ad hoc polysemies do not pass the co-predication and anaphoric tests. So, even these tests are not foolproof, being liable to give false negatives. Still, they do suggest that the distinction between homophony and polysemy is one with a difference. Numerical subscripts will not allow the truth-conditional semanticist to answer the challenge from polysemy.

4.3.2 POLYSEMY IS NOT INDEXICALITY

Another strategy would be to appeal to indexicality, or to exploit in some other way the resources already afforded by the standard truth-conditional approach when it comes to context-sensitivity. For example, we might consider positing variables in the syntax or semantics of polysemous expressions, variables that would be assigned values by the assignment function *g* that is determined in a context *c*. However, this seems to be unmotivated. Formally, we treat pronouns and other deictic expressions as variables because such expressions can occur either bound (as in 'Every boy was tired when he arrived') or free ('He arrived'). To explain the possibility of bound readings, we appeal to variables. And, since we already treat bound pronouns as variables in order to explain judgments about binding, it is reasonable to treat free occurrences in the same way, allowing the context to determine a mapping from indices to values. And this treatment of free pronouns is not merely a convenient result of our treatment of bound occurrences. In effect, speakers use pronouns and demonstratives and the like (when unbound) to refer to jointly salient things, and the assignment function *g* therefore corresponds to something like a ranked ordering of discourse referents in terms of relative salience. However, by contrast, there is no correlate of the bound/free distinction with respect to polysemous terms. And, in conversation, we do not track the relative salience of different resolutions of polysemy in the way that we track the relative salience of discourse referents (Lewis 1979). So, we have very little reason to hypothesize that polysemy is covert indexicality.

A related strategy, then, might be to enrich the index of evaluation, adding

an additional parameter *p* that in effect resolves polysemy. However, there is no single feature or set of features (other than the speaker's intentions) relative to which polysemy gets resolved. The sense of a particular occurrence of '$bank_1$' is not resolved by considerations about who is speaking, or where, or when, or to whom; again, resolution of polysemy is not driven by consideration of salient contextual features like those that are included in formally defined contexts. What determines the resolution of polysemy on a given occasion appears to be the speaker's intentions. Therefore, in effect, to appeal to a polysemy resolution parameter *p* in the index of evaluation is to say that (17) uniquely determines a proposition relative to the speaker's intentions (or relative to someone who knows them).

(17) The $bank_1$ is collapsing.

Something along these lines has been proposed with regard to demonstratives (King 2012). However, it is one thing to suggest that *speaker* reference is determined by speaker intentions; it is quite another thing to include speaker intentions in the index of evaluation relative to which a *sentence* is mapped to a proposition. This latter strategy seems to be incompatible with serious semantic theorizing. What would put principled restrictions on the explanatory scope of this kind of semantic theory? What need would there be for *g*, if it suffices to say that reference is determined by speaker intentions? In short, then, we cannot assimilate polysemy resolution to indexicality, neither by appealing to covert variables nor by appealing to an extra parameter in the index.

4.3.3 LITERALISM, MINIMALISM, CHIMERAS, AND POINTERS

To sum up so far, within the truth-conditional approach to semantic theory, polysemy cannot be assimilated to homophony; and polysemy cannot be assimilated to indexicality, either. What options are left to the truth-conditional semanticist? There are, in principle, at least three to consider before one abandons truth-conditional semantics for an alternative approach. The first would be to posit one sense of a polysemous term as privileged, treating any other senses as the result of pragmatic processes of modulation; call this 'literalism.' (Versions of literalism for at least some categories of polysemy have been defended, for example, in Copestake & Briscoe 1992; Asher 2011; cf. Falkum 2011 for a Relevance Theoretic version of the literalist approach.) The second would be to appeal to a single, highly general sense that subsumes the various specific senses with which polysemous terms are used; call this 'minimalism' (e.g., Borg 2005; Cappelen & Lepore 2005) A third strategy would be to appeal to "dot objects," chimeric combinations of the various senses of a polysemous term; call this 'chimeric semantics' (e.g., Langacker 1991; Asher 2011).

Literalism is at least initially promising with regard to some non-inherent, irregular polysemies (Vicente 2018). It is not unreasonable to suggest that there is some privileged, literal sense of 'line,' from which the others are pragmatically derived via processes of modulation (cf. the various uses of terms for body parts, such as 'mouth,' 'eye, 'hand;' or the uses of perceptual terms like 'red' or 'warm' to describe emotions and personality traits; or uses of emotion terms like 'sad' to describe music). The same goes for 'chicken' or 'lamb': it is in fact intuitively

plausible to suggest that the sense related to countable organisms is the term's privileged, literal meaning, the sense related to meat being derived therefrom via pragmatic mechanisms. Literalism is perfectly compatible with truth-conditional semantics: the proposition uniquely determined by a sentence including a polysemous term with a privileged sense is going to be the proposition compatible with that sense, not one of the pragmatically derived senses. (See Falkum & Vicente 2015 for an overview of literalism, along with many citations.)

However, as a strategy for accommodating polysemy, literalism has a limited scope. Take the regular and inherent polysemy of '$bank_1$,' again. Which is its privileged or literal sense: the one having to do with institutions, or the one having to do with buildings? There seems to be no principled answer to this question. We do not have intuitions about which sense of 'bank' is primary, in contrast with 'chicken' (Vicente 2018: 959-960). We can stipulate that one of the senses of '$bank_1$' (and 'book,' and so on, for all inherently and regularly polysemous terms) is its privileged sense. But any such stipulation would appear to be arbitrary and ad hoc, with apparently no independent motivation aside from the desire to keep truth-conditions in semantic theory.

The minimalist approach is similarly limited in its ability to handle polysemy within a truth-conditional framework. For the minimalist, the semantic output given a single word is a highly general property, one that subsumes the various more specific properties that speakers manage to predicate in using the word (Borg 2004; Cappelen & Lepore 2005). So, at least in principle, it seems that the minimalist avoids the problem of having to choose one of the specific

senses of a regular polyseme as its privileged meaning. However, there does not seem to be a single, highly general property that is had by both physical buildings and financial institutions. This is not to say there are no such things as minimalist properties. It is plausible enough to suggest that there is some highly general property that is the privileged, albeit very under-specific sense of 'blue.' The minimalist can acknowledge that things are blue in various specific ways (lighter or darker, on the outside or the inside, in whole or in part), but nonetheless maintain that these different ways of being blue are all in some way subsumed by the highly general property of being blue. However, this is implausible with regard to inherent polysemy. It is implausible to insist that there is some highly general property that being a physical building and being a financial institution are specific instantiations of. So, while the minimalist avoids the trouble of having to choose one such sense as the privileged meaning of '$bank_1$,' the minimalist does so in a way that leaves us without any candidate for the minimalist meaning of '$bank_1$.' While the minimalist might be tempted at this point to appeal to homophony and deny that '$bank_1$' is polysemous, or to appeal to indexicality, we have already seen good reason to reject such a strategy.

This leaves chimeric semantics as the only viable strategy for dealing with inherent polysemy on a truth-conditional framework (cf. Vicente 2018). The chimeric semanticist appeals to the dot objects of Pustejovsky 1995. Dot objects are a special semantic type; '$bank_1$' is not of the type information or physical object, but rather of the type information•physical object. The two specific

types are two "aspects" of the dot object, which can be "highlighted" differentially, as in (27) and (28).

(27) My favorite $bank_1$ is the one with the revolving doors.

(28) I like a $bank_1$ that keeps my savings off Wall Street.

But, as shown above, inherently polysemous terms pass the co-predication test, with (24) being judged acceptable.

(24) The $bank_1$ has a weathered façade but a good reputation.

In this way, a chimeric account of inherently polysemous terms is compatible with a truth-conditional semantics. On such an account, any sentence including '$bank_1$' will uniquely determine, relative to an index and a rule of composition, a proposition that includes a dot object as a constituent. Such propositions might strike some theorists as ontologically profligate, but I shall press no such objection here. And, of course, the propositions that speakers directly assert typically involve only one aspect of a dot object (as in a typical utterance of (17)), and not the complete dot object itself. But this is no threat to truth-conditional semantics, an approach that is already willing to acknowledge many other divergences between speaker and sentence meanings. All '$bank_1$'-containing sentences uniquely determine (relative to an index of evaluation and rule of composition) a proposition that includes a dot object as constituent. On the way to attributions of speaker meaning, hearers frequently filter out, based on extra-linguistic contextual considerations, parts of the dot objects that are paired by their semantic competence with the polysemous words they hear.

So far, my goal has been to lay out the options available to a truth-conditional semanticist in the face of the challenge from polysemy, and to identify those strategies that seem most promising. Inherent, regular polysemy is not amenable to treatment via any other strategy, and so chimeric semantics seems like the truth-conditional semanticist's best option.

Of course, an alternative to chimeric truth-conditional semantics is a semantics without truth-conditions. Pointer semantics is one such alternative, motivated in part by consideration of polysemy. A pointer semantics pairs sentences with "pointers" to conceptual space, instructions to fetch a concept, or constraints on expressible contents, where said constraints are in general too weak to determine a propositional content uniquely, given just an assignment of values to variables and an index of evaluation (Carston 2002b, 2012; Pietroski 2006, 2008, 2010, 2018; Harris 2020.). A constraint semantics maps 'Bob speaks' not to a context-relative propositional content, but rather to a constraint on the propositions that a speaker could assert in uttering that sentence and, accordingly, on the propositions that a hearer could take a speaker who utters that sentence to have asserted. And a lexical item like 'Bob' is mapped to neither a referent nor a representation of a referent (i.e., a sense), but rather to a constraint on who or what a speaker could refer to with that expression (or a constraint on the concepts that a hearer could access upon hearing that expression).

Pointer or constraint semantics is well suited to accommodate any kind of polysemy. On one contemporary version (Pietroski 2018), couched within a heavily mentalistic (albeit competence-oriented, not performance-oriented)

semantics, the linguistically encoded meaning of a lexical item is an instruction to fetch a concept at a mental address, while the meaning of a phrasal expression is a complex instruction to fetch multiple concepts and compose them in certain ways. The meaning of '$bank_1$' is thus an instruction to fetch a concept at a mental address—an address at which multiple distinct but related concepts are located. (A homophonous expression such as '$bank_2$' would encode a different instruction.) One of these concepts must then be selected from the indicated address and then "passed along" for further processing on the way to an attribution of speaker meaning. As I discuss below, the differences between the chimeric and pointer semantics are fairly subtle, and especially so when we focus on processing.

In this section, I have argued that (i) the standard, truth-conditional approach in formal semantics faces a challenge with regard to polysemy and (ii) that regular polysemy in particular forces us to appeal to chimeras in order to keep truth-conditions in semantics. In the next section, I reassume the cognitivist perspective within which I have approached semantics and pragmatics throughout most of this dissertation. However, the discussion of formal semantics is not an idle sidecar. Polysemy raises a parallel challenge within the cognitivist framework. And, as I will show, the same options are available for the cognitivist who, having recognized this challenge, would nevertheless like to maintain the hypothesis that the workings of the semantic input system eventuate in something truth-conditional (i.e., a thought).

4.4 POLYSEMY FROM A COGNITIVIST PERSPECTIVE

Within a cognitivist framework, polysemy raises a similar challenge. In short, a common idea among cognitivists is that a word is mapped by the semantic input system to something conceptual (chapter 1). Given a rule of composition for concepts, and given a sentence as input, the workings of the semantic system might then eventuate in something truth-evaluable (i.e., a thought). However, polysemy threatens this common idea, since it is not immediately clear what concept 'bank$_1$' or 'book' is mapped to. As I argue, the choice again comes down to (a cognitivized version of) chimeras or pointers. Agustin Vicente (2018) argues in favor of chimeras; but his argument rules out only literalism and minimalism, not constraint semantics. What is needed, then, is some way to adjudicate between the chimera and pointer approaches to polysemy within a cognitivist framework. The cognitivist is answerable to data from psycholinguistic experimentation. Accordingly, I will articulate predictions that might differentiate between these two theories, which have so far not been tested against one another. First, however, some unpromising theories need to be set aside within a cognitivist framework.

4.4.1 SOME DEAD ENDS

Many of the options that were briefly considered and then rejected above in relation to the orthodox, truth-conditional approach to semantics can be all the more easily set aside by the cognitivist. Some experimental results reinforce the lesson of the co-predication and anaphora tests (for overview and discussion, see

Eddington & Tokowicz 2015; cf. Falkum & Vicente 2015, Vicente 2018, Carston 2020).

The different senses of a polysemous word prime each other, while the different senses of a pair of homophones show effects not of priming, but rather of competition (Klepousniotou et al. 2008). And words with multiple related senses tend to be responded to faster than words with fewer senses (Azuma & van Orden 1997; Rodd et al. 2002; Klepousniotou & Baum 2007). If all senses were represented distinctly in the lexicon, then having more senses would lead to slower responses (if all senses are accessed) or make no difference (if only the relevant sense is accessed). Also, there is a response bias for more frequent senses of homonyms, but not polysemes, (Duffy et al. 1988). Finally, hearers/readers make an immediate decision about homophony resolution (usually within 250ms), while holding off for a while for polysemes (Frazier & Rayner 1990).

The apparent lesson of the studies reviewed so far is reinforced by experimental work involving magnetoencephalography (MEG). The MEG 350 component is thought to reflect initial stages of lexical activation and is generated in the left superior temporal cortex 300-400ms after visual word onset. Evidence suggests that this component reflects morphological root access (Pylkännen and Marantz 2003; Fiorentino and Poeppel 2004; Embick et al. 2001). Suppose we take for granted one of the lessons of Klepousniotou et al. 2008: the senses of homophones compete with one another for activation. Priming and competition can be observed by way of comparing peak latencies of the MEG 350 component. An earlier peak relative to controls indicates priming, while a delay indicates

competition. These linking assumptions were used along with the following experimental conditions to test whether polysemy patterns together with homophony in showing effects of competition among senses, or if instead the senses of a polysemous term facilitate or prime one another.

	PRIME	CONTROL	TARGET
HOMOPHONY	river bank	salty dish	savings bank
POLYSEMY	lined paper	military post	liberal paper
SEMANTIC	lined paper	clock tick	monthly magazine

If the view that the senses of a polysemous term are stored separately in the mental lexicon is correct (i.e., if polysemy is in fact a case of homophony), then we should observe a greater delay in peak latency from M350 sources in the polysemous conditions as compared to the semantic relatedness conditions. However, we do not observe such a delay. Instead, we find similar priming effects in the polysemy conditions and the "semantic" conditions, where the latter are the paradigmatic priming conditions. The homophony condition, by contrast, did show a greater delay in its peak latency from the M350 source, as compared to polysemy and "semantic" conditions. MEG recording was paired with a behavioral task requiring word identification. The behavioral observations support the MEG observations, with delayed responses in the homophonous conditions as compared to the polysemy and "semantic" conditions, with the latter two conditions giving rise to similar timing profiles (Pylkkänen et al. 2006).

These results, together with the more traditional linguistic tests described above, rule out the strategy of treating polysemy as homophony. The linguistic mind stores the senses of polysemes in a different way than the senses of

homophones, and the communicating mind processes polysemous and homophonous input in different ways. Granted, the data in this regard are not unequivocal, and it can be argued that at least some putative cases of polysemy should in fact be reclassified as homophony (Klein & Murphy 2001; Foraker & Murphy 2012) Still, the scope of this strategy is limited, since the very same studies that suggest that some putative examples of polysemy should be reclassified as homophony also reinforce the conclusion that there are some genuine examples of polysemy.

Assimilation of polysemy to indexicality is likewise not a viable strategy for cognitivists. From the perspective of processing, it is especially unclear how appeal to a polysemy resolution parameter *p* differs from appeal to general human intelligence. Consider (17) again.

(17) The $bank_1$ is collapsing.

Speakers usually intend one resolution of '$bank_1$' when uttering a sentence like (17). And the audience can sometimes figure out what that intention is. In resolving the polysemy of '$bank_1$', hearers may often rely on co-text ('savings bank' versus 'tall bank'). But, often enough, they cannot and/or need not rely on co-text. In the right conversational context, a speaker can readily communicate a proposition about a building by way of an utterance of (17). This is because polysemy resolution can in principle be guided by any piece of information locally available (linguistic or not) to a hearer. Hearers are generally expected to be rational and cooperative. Crucially, hearers are generally expected to rely on more than co-text in attributing speaker meanings, including facts about what is

salient in the discourse context, but also expectations about what the speaker is likely to intend, and general considerations about what combinations of predicates seem sensible—considerations that, again, do not pertain to anything that we could consider a contextual feature in a formal semantic theory.

Within a cognitivist framework, appeal to the assignment function *g* in a formal semantics can plausibly be taken to correspond to something attentional. We use deictic expressions to refer to jointly salient things. Accordingly, speakers and hearers keep track of the relative salience of various discourse referents (Lewis 1979), as is made clear when we observe the care with which speakers select their referring expressions ('John' versus 'the man over there' versus 'him' versus etc.), and the speed with which hearers resolve the reference of such expressions. However, resolution of polysemy does not seem to be a matter of attentional salience. We do not, in speaking to one another, keep track of the relative salience of different predicates (or resolutions of polysemy). Again, speakers usually intend one resolution of 'bank_1' when uttering a sentence like (17). And the audience can sometimes figure out what that intention is, based on various considerations, including recognition of the word(s) uttered. The question, then, is *what* the hearer is provided with *given only* word identification: a privileged, specific concept; a highly general or minimalist concept; a chimeric assemblage of concepts; or something entirely non-conceptual, like a pointer to a region in conceptual space?

Literalist and minimalist approaches to the processing of polysemy run into familiar problems. For the literalist, given word recognition, a polyseme is

context-invariantly mapped to a single privileged concept from which other, non-privileged concepts can be pragmatically derived. For the minimalist, given word recognition, a polyseme is context-invariantly mapped to a highly general concept from which more specific concepts can be pragmatically derived. Each of these strategies is at least initially plausible when we consider certain merely regular ('chicken') or irregular polysemies ('line'). However, there seems to be no single concept that is the privileged meaning of 'bank,' nor does there appear to be a highly general concept that subsumes the more specific concepts, INSTITUTION and BUILDING. At best, the literalist approach has only a very limited scope within a cognitivist semantic theory.

This leaves chimeras as the only option if we want the semantic input system to eventuate in a truth-evaluable thought (albeit a chimerical one). The alternative, again, would be to abandon truth-evaluability in a cognitivist semantics and take a different approach altogether. And, in many ways, the constraint-based approach (what I have been calling 'pointer semantics') is well suited to cognitivism. Many cognitivists have therefore embraced constraint or pointer semantics (see, e.g., Carston 2012; Harris 2020); though, as I showed above, sometimes as a result of overly hasty arguments from encapsulation (as in Harris 2020; see section 4.1.2 above). However, it has also been argued that, on the contrary, consideration of polysemy favors chimeras over pointers (Vicente 2018). In the next section, I review this second line of argument, and show why it, too, is unsuccessful.

4.4.2 VICENTE'S ARGUMENT FOR CHIMERIC SEMANTICS

Vicente (2018) offers a sequence of arguments meant to show that the phenomenon of polysemy lends favor to chimeric ('overspecific,' in Vicente's terms) semantics, at least for at least some polysemous words. The alternatives considered are literalism, on the one hand, and an underspecified semantics, on the other. By 'underspecific,' Vicente intends to capture both what I have called 'minimalism' and what I have called 'constraint semantics.' Vicente argues in favor of chimeric semantics for at least a restricted class of polysemous lexical items—namely, for so-called inherent polysemes—in a few steps.

First, Vicente rehearses the experimental evidence in support of the polysemy-homophony distinction, as just described above. On the basis of this record, we cannot assimilate polysemy to homophony within a cognitivist framework (Vicente 2018: 953). Vicente furthermore argues that "[r]ecent psycholinguistics tends to favor underspecification and overspecification accounts" over a literalist semantics (Vicente 2018: 952). Here, Vicente relies on just the sorts of considerations mentioned above in setting aside literalism as an account of polysemy. For so-called inherently polysemous words, such as '$bank_1$,' it stretches credulity to suggest that one of the conventionalized senses is privileged over the others.

Third, Vicente argues that there is no "abstract representation that encompasses all the senses" of logical and regular polysemes, ruling out an underspecified approach (2018: 959-960). On Vicente's characterization, underspecific semantics maps a word to a 'common core.' So, e.g., the common

core of 'cut' might be "change of state in which an entity which exemplifies some kind of connectedness undergoes a process of controlled disconnection" (Spalek 2015; Vicente 2018: 952). In contrast with 'cut,' Vicente points out, we cannot specify a common core for 'bank$_1$,' one that encompasses the institutional and physical senses. The only thing that can encompass the various senses of inherent polysemes is a list, or chimeric assemblage of concepts. According to Vicente, this leaves chimeric semantics, an over-specific semantics, as the account best suited to explanation of polysemy.

The third, crucial step in Vicente's argument is flawed. In short, this argument neglects the distinction between two versions of underspecified semantics: the abstract representations that are the standing meanings of polysemous lexical items can either be conceptual, or not. If they are conceptual, we end up with a minimalist version underspecified semantics: the semantic output given an open-class word is conceptual, but highly under-specific with regard to the concepts that hearers use that word to express. If words are mapped to something that is under-specific but not conceptual, we end up with a version of pointer semantics. This is a distinction to which Vicente does briefly allude (2018: 950). However, his argument against underspecified semantics only works as an argument against the conceptual version. It is unsuccessful with regard to the non-conceptual version of an underspecified semantics—i.e., with regard to pointer semantics.

Again, Vicente's objection to underspecified semantics is that there is no common conceptual core that is shared by the senses of an inherently polysemous

word like ‘bank$_1$.’ There is no non-disjunctive conceptual entity that could encompass both INSTITUTION and LOCATION without also encompassing any number of other concepts, not among the candidate meanings for a given use of ‘bank$_1$.’ In searching for a common conceptual core for an inherently polysemous word like ‘bank$_1$,’ we find nothing that could do the job. This much is correct, I agree. But Vicente is wrong to conclude that this observation rules out an under-specified semantics for inherently polysemous words; it only rules out a conceptual version of under-specific semantics. The non-conceptual alternative has it that inherently polysemous words (perhaps along with other kinds of polysemy) encode not a conceptual core, but rather something like a pointer to a region of conceptual space. In pointing to a region of physical space, my pointing gesture does not *encompass* the points that occupy that region. Likewise, a pointer to a vague region of conceptual space does not encompass the concepts in the region indicated. An instruction does not encompass the varying ways in which it can be carried out. A constraint does not encompass the ways in which it can be satisfied. Pointers, instructions, constraints—these are not conceptual common cores, or conceptual entities of any kind: the relation between a pointer and a concept differs in kind from the relation between a common core and a specific application of it, and only the latter is accurately described as a relation of encompassing.

The constraint semanticist pairs lexical items not with abstract conceptual representations, but with something under-specific and non-conceptual: constraints on (or instructions for, or pointers toward) concept retrieval. On this

view, the fact that there is no "abstract representation that encompasses all the senses" (o.c.) of a polyseme is no problem: constraints do not in any sense encompass the various ways in which they can be satisfied. While Vicente's argument supports the conclusion that chimeras are the best option if we want a truth-conditional cognitivist semantics, the considerations he raises simply do not rule out a non-truth-conditional, non-conceptual, constraint-based approach.

4.4.3 DIFFERENTIATING CONSTRAINTS AND CHIMERAS

So, within a cognitivist framework, we are left with two options when it comes to regular polysemy. We can go with chimeras, or we can go with pointers as the immediate output of a semantic system that maps word-forms to mental entities. To decide between these options, what we need are considerations and predictions that differentiate chimeric and pointer semantics. Here is one promising line of thought. The set of concepts linguistically associated with a lexical item is a discrete set of privileged concepts on chimeric semantics, and a non-discrete and penumbral cluster of concepts on constraint semantics. This distinction seems to involve a real difference, one that is especially clear when we consider how each view explains ad hoc polysemy.

On a pointer semantics, words are not intrinsically associated with anything conceptual. Rather, a word is a "pointer to conceptual space" (Carston 2012). Appealing to yet another metaphor, we might say that a word is a flashlight: it shines a light on some concepts that occupy a certain region or neighborhood. The area indicated includes some number of individual concepts. On the way to an

attribution of speaker meaning, the hearer retrieves one concept, among the many illuminated by the word, that fits with the top-down, pragmatic pressures of the speech situation. Furthermore, because some of the concepts in the illuminated region are near its boundary, and because this boundary is penumbral or vague, the concept that a hearer ends up selecting and passing along for downstream processing (and, ultimately, attribution of speaker meaning) is sometimes within the penumbra of the indicated region. In some cases, ad hoc polysemy is thus a matter of fortuitous selection of a concept along the boundary of the indicated region in conceptual space. Additionally, all that is context-invariantly afforded by the linguistic input itself (i.e., the word) is activation of the pointer or instruction, not activation of the specific concepts that are in the indicated region. Having recognized a word/pointer, one particular concept in the indicated region might receive increased activation (in effect, "jumping out" visually from the scene illuminated by the flashlight) as a result of top-down pressures of extra-linguistic context. In other words, the pointer view does not predict that the various senses of a polysemous word are context-invariantly activated whenever said word is input to the semantic system.[36]

According to chimeric semantics, a word provides a discrete list of concepts, and the hearer selects one that fits with the top-down, pragmatic pressures of the speech situation. A word maps context-invariantly to a discrete list of privileged

36 These metaphors (instructions, pointers, flashlights) crucially involve some appeal to, in effect, an internal faculty of attention, such that the semantic processor is able to attend (or eventuate in instructions for some other system to attend) to different regions in conceptual space. As far as I know, cognitivist versions of this pointer-based approach have not appealed to any of the specific models of attention available in cognitive science more generally. This is another respect in which the view stands in need for more development.

concepts. Top-down pressures may influence selection, but the chimeric view requires that every part of the list is activated given word recognition. So, in contrast with the pointer view, the chimeric view does indeed predict, across all contexts of use, increased activation of all the chimera's parts given recognition of a polysemous word. Furthermore, on the chimeric view, there is no correlate of the penumbral vagueness appealed to by the constraint semanticist. Ad hoc polysemy, on the chimeric view, is always a matter of inferring that the meant concept is not among those made available by the word uttered, and then inferring the meant concept. So, in contrast with the pointer view, the chimeric view predicts that distinct mechanisms are involved in the resolution of conventionalized versus ad hoc polysemies.

To sum up, the chimeric semanticist predicts that, upon recognizing a given word, we should find evidence for activation of all the conceptual parts of the chimera to which that word is mapped, regardless of whether they are all contextually supported. That is, we should see evidence that a privileged and discrete list of concepts receives increased activation in a way that is stubborn against top-down pragmatic effects. Each concept in the chimera must context-invariantly be made available given word recognition, although (usually) only one is selected on the way to an attribution of speaker meaning. On the other hand, the pointer semanticist does not predict that unmeant concepts get stubbornly activated in this way. In addition, the chimeric semanticist predicts that distinct mechanisms are involved in resolution of conventionalized versus ad hoc polysemies, while the pointer semanticist leaves open the possibility that (at least,

in some cases) the two kinds of indeterminacy might be resolved by the same mechanism.

These differences in turn suggest a way of adjudicating among these alternative cognitivist conceptions of word meaning. Systematic absence of evidence for the kind of context-invariant stubbornness described just above would be problematic for the chimeric theory, which predicts such stubbornness, and favorable to the pointer theory, which could easily explain the apparent absence of stubbornness. On the other hand, evidence that word recognition does context-invariantly lead to activation of a discrete list of privileged concepts would support the chimera theory and leave the pointer theorist in need of an explanation for this stubbornness, which is not predicted by the pointer theory. In short, the pointer or constraint theorist would have to say that, when we observe stubborn activation of unmeant concepts, this is always the result of top-down effects, whereas, for the chimera theorist, such stubborn activation is the inevitable, bottom-up result of word recognition.

As I have argued, the extant arguments concerning polysemy do not adjudicate between chimeras and constraints. While the available experimental record convincingly shows that homophony and polysemy are distinct phenomena, with the senses of homophones being stored and processed differently than the senses of polysemes, currently available data do not inform the choice between lists and pointers as the output of a semantic process given a regular polyseme as input. We do not currently have any psycholinguistic evidence, along the lines sketched just now, that would allow us to adjudicate

between chimeras and pointers (as noted in Frisson 2009; MacGregor 2015; Vicente 2018). However, as we look to future experimental studies of polysemy in hopes that they might inform the choice between pointers and chimeras, there are a few things to keep in mind, as I discuss in the rest of this chapter.

The pointer semanticist faces the familiar challenge of maintaining a distinction between the concepts that are possible meanings of a given word, and the concepts that are merely associated with that word. Generally, we assume that the former is a proper subset of the latter; in other words, some but not all of the concepts associated with a word are among those that a speaker can mean *in* using the word. Regarding '$bank_1$,' for example, we assume that INSTITUTION is one of its possible meaning, but not TELLER. Although, of course, TELLER is associated with the word '$bank_1$' and with whatever concepts are among its possible meanings, TELLER is not among the possible meanings of '$bank_1$.'

So, not all conceptual associates are possible meanings of a word. Or, to put the point another way, not all associations are polysemies. The chimeric semanticist can easily preserve this distinction, since each word is mapped to a discrete list of privileged meanings. Not all the concepts associated with '$bank_1$' are on the list, or part of the chimera. Why is INSTITUTION something that a speaker can mean in uttering '$bank_1$' while TELLER (or MONEY, or etc.) is not? The chimeric semanticist can quickly answer this question by appeal to the basic apparatus of her account: because INSTITUTION is one part of the chimera, but TELLER is not. So, it is a virtue of the chimeric view that it readily affords the means to preserve the distinction between associations and meanings. (Of course,

the chimeric semanticist then needs to offer an account of how chimeras come to exist and how they come to have the parts that they have. But the pointer semanticist faces a similar problem: how did the concepts that are in certain regions of conceptual space end up in the same neighborhood, and how did that neighborhood come to be linked to a particular linguistic item?)

The pointer semanticist, by contrast, cannot capture the meanings/associations distinction as straightforwardly. Word recognition guarantees retrieval of a pointer or instruction, which in turn activates (perhaps, to some degree, depending on top-down contextual pressures) some number of concepts; the process of selection then involves fetching one of these concepts on the way to an attribution of speaker meaning. Again, on this view, it is possible that, in some cases, the concept selected is selected fortuitously, as a result of top-down pressures. And a concept can be selected from the penumbral boundary of the indicated region. So, a pointer semantics on its own gives us no principled reason to deny that TELLER is among the possible (directly assertable) meanings of 'bank_1.' In other words, the pointer semanticist's theory does not directly entail a principled distinction between meanings and associations. And the distinction between ad hoc polysemy, on the one hand, and metaphor, on the other, also seems to be blurred.

So, what might be seen as a virtue of pointer semantics—i.e., its ability, by appeal to a vague region of conceptual space instead of a discrete list of privileged concepts, to give a unified account of the mechanisms underlying resolution of ad hoc and conventionalized polysemy—is also a potential vice. Again, that vice is

the loss of a principled distinction between meanings and associations, and perhaps also that between polysemy and metaphor.

The pointer semanticist might then simply respond by suggesting that the indicated region of conceptual space must have a sharp boundary, and the same boundary across contexts of use. However, the pointer view then becomes virtually indistinguishable from the chimeric view. The chimeric view says that, for a polysemous open-class word *e*, there is a discrete list of concepts $c_1 \ldots c_n$, such that all of $c_1 \ldots c_n$ are activated whenever *e* is input to the semantic word processor. On the proposed modification of the pointer view, it says that, for a polysemous word *e*, there is a region of conceptual space including some specific number of concepts $c_1 \ldots c_n$ such that, whenever *e* is recognized, the region of conceptual space including $c_1 \ldots c_n$ is context-invariantly activated. In other words, the chimeric semanticist appeals to the word and the list of concepts; the pointer semanticist (at least, one who abandons the idea that the indicated regions in conceptual space are vague) appeals to the word, its corresponding list of concepts, and a pointer that mediates between the word and the list. However, when the list of concepts is discrete, it is unclear what word the pointer does, over and above the work already done by the pairing of a word and a list of concepts.

Hence, two main predictions differentiate the pointer and the constraint view. The chimeric view predicts that we should see evidence that discrete list of privileged concepts receives increased activation in a way that is stubborn against top-down pragmatic effects. Each concept in the chimera must context-invariantly be made available given word recognition, although (usually) only one is selected

on the way to an attribution of speaker meaning. On the other hand, the pointer semanticist does not predict that unmeant concepts get stubbornly activated in this way. In addition, the chimeric semanticist predicts that distinct mechanisms are involved in resolution of conventionalized versus ad hoc polysemies, while the pointer semanticist leaves open the possibility that (at least, in some cases) the two kinds of indeterminacy might be resolved by the same mechanism.

These differentiating predictions afford a promising route for adjudicating among pointers and constraints. But they also raise new challenges. As argued here, one virtue of pointer semantics is that it allows (by appealing to vague, non-discrete regions of conceptual space, as opposed to discrete lists of privileged concepts) some ad hoc polysemies to be resolved in the same way that regularized polysemies are resolved—by the same semantic system, as opposed to a downstream, post-semantic inferential process. But this virtue comes with a problem: the blurring of pre-theoretically important distinctions, such as that between meanings and associations, and between ad hoc polysemy and metaphor. As cognitivists, however, we should be prepared to find that, in fact, these theoretical distinctions *are* blurred in the hearer's mind. Perhaps, ultimately, the experimental record will convince us to abandon the hope of preserving a principled distinction between semantics and pragmatics in an account of the moment-by-moment processing of linguistic input. I turn to such "big picture" questions in the next and final chapter.

Chapter 5: Caveats and conclusions

I have adopted here a cognitivist perspective that focuses on the processes and mechanisms that are involved in the moment-by-moment interpretation of utterances. Among cognitivists, it is commonly thought that the distinction between semantic and pragmatic aspects of linguistic understanding is, in effect, meant to explain how hearers solve a frame problem that arises during utterance interpretation. The hearer's job is to attribute communicative intentions to a speaker, or to infer what messages the speaker means (with varying degrees of directness) to communicate. Any piece of information is potentially relevant to figuring out what a speaker means, even in the relatively direct sense. The Fodor-style encapsulation hypothesis for semantics posits a linguistic decoding system that includes an encapsulated semantic component, one that retrieves and combines mental representations in accord with the structure of the linguistic input.

By hypothesis, the semantic component of the decoding system is insusceptible to top-down, extra-modular influences. Most of the information available in the mind is incapable of influencing the operations of the semantic processor. This Fodor-like proposal regarding semantics is attractive, since it promises a part of the explanation for how hearers solve the frame problem that arises for utterance interpretation. While any piece of information is potentially relevant to inferences about speaker meaning, these inferential processes are constrained by the affordance of the encapsulated linguistic system. The semantic

decoding system operates over linguistic input without being susceptible to the top-down influence of almost all the information available in the hearer's mind (with the exception only of the information made available by the input, or included in the system's proprietary database). And the deliverance of this encapsulated semantic processor, given a piece of linguistic input, serves as evidence that guides the inferential processes that result in attributions of speaker meaning. Not only does the linguistic input provide evidence that constrains belief fixation regarding speaker meaning, but the linguistically afforded evidence is itself fixed in a perception-like way, and not as a result of isotropic inference.

However, as discussed in chapter 2 above, the case in favor of this Fodor-style view of semantic processing is not especially strong. In rapport with parallel debates about the encapsulation of vision, I have offered objections to some of the standard arguments for encapsulated semantics. For one thing, the fact that the encapsulation of semantic processing would preserve a principled distinction between semantic and pragmatic information in a science of utterance interpretation does not give us very good reason to endorse the encapsulation hypothesis. It is an empirical hypothesis to say that there is a semantic decoding system that operates over linguistic input without being susceptible to top-down, extra-modular effects. Approaching semantics in this processing-oriented way entails that the cognitivist is answerable to the methods and results of psychology. Accordingly, in chapter 3 above, I reviewed several experimental observations that at least initially appear to be pertinent to assessment of the hypothesis that semantic operations are encapsulated. As I have shown, there is quite a lot of data

to suggest very early effects of top-down influences on the process of content resolution, or assigning a concept to an identified word. As I have also demonstrated, there are principled ways in which the Fodor-style hypothesis for content resolution can be maintained in the face of this empirical challenge. Looking again to vision, I pointed to what I take to be especially compelling empirical arguments against Fodor-style encapsulation in that domain. There is evidence to suggest that activity in the brain regions that are known to be responsible for early visual processing show top-down, extra-modular effects of general cognition (see Ogilvie & Carruthers 2016). I have proposed applying this strategy in the case of semantic processing, looking to some recent neuroanatomical models of semantic processing that might provide starting points. Future work in psycholinguistics might pursue this strategy further.

It has been argued that certain claims about the outputs of semantic processing follow from the Fodor-style encapsulation hypothesis for semantics (Borg 2004, Harris 2020). However, as I showed in chapter 4, these arguments are unsuccessful. The encapsulation of a certain stage of content resolution is compatible with several hypotheses about what it might yield as output given a word as input: a highly general concept, a list of concepts, or perhaps a pointer to conceptual space. As I have argued, consideration of polysemy does seem to shorten the list of options to the second or third of these. A semantic processor that is encapsulated in the Fodor-like way cannot yield the concept FOLIO given 'book' as input. However, as I discuss, no currently available arguments or experimental results adjudicate among the list (or chimera) theory, on the one

hand, and the pointer theory, on the other hand. Accordingly, I sketched at the end of chapter 4 the beginnings of a proposal regarding how to adjudicate among these two views. As acknowledged, however, much of the details of the chimeric and pointer-based accounts are metaphorical, and such metaphors would need to be cashed out with explicit models. Though I have offered no such models here, I have distinguished the two accounts both conceptually and in terms of the predictions they make.

If the Fodor-style hypothesis for semantic processing turns out to be false, then what should we conclude regarding a mentalist linguistics? As I have argued, cognitivists might consider alternatives to the notion of Fodor-style modularity in order to preserve a principled distinction between semantics and pragmatics in a science of utterance interpretation. One idea considered here is to appeal to massively modular models of the mind, which allow for top-down, extra-modular effects on modular processes (e.g., Carruthers 2006). This non-Fodorian conception of modularity for semantic processing does not entail the loss of the semantics-pragmatics distinction in a science of utterance interpretation. If the semantic decoding system retrieves and composes word meanings in such a way that it is blind to most of the information available in the mind during its operations, this would preserve a version of the distinction in a science of utterance interpretation. That no piece of information is in principle incapable of influencing a certain processing system does not entail that its operations are a "whole brain," inferential affair.

That said, the extent to which the semantics-pragmatics distinction is preserved in the performance of massively modular mind will depend on the degree to which the operations of the semantic processing system are penetrable by top-down influences. Suppose we were convinced that the processes involved in what cognitivists call 'semantic decoding' (retrieving and composing concepts in accord with linguistic input) were regularly shot through with such influences. Ultimately, as briefly mentioned above (section 3.3.2), we might conclude that there is no principled distinction between semantics and pragmatics in a science of utterance interpretation. From the perspective of performance or use, we might conclude that there are some narrowly linguistic processes of phonetic and syntactic processing that are encapsulated (at least to a relatively high degree, if not in the Fodor-style sense), but that the rest is a matter of mostly isotropic inference plus heuristics.

We have good reason to think that knowledge of language is distinct from knowledge of the world more generally, and accordingly that to know the meaning of 'dog' is not just a matter of knowing all sorts of facts about dogs. However, as I have argued, we cannot simply assume that such a distinction will be recapitulated in the processes and mechanisms that are involved in language use and the moment-by-moment interpretation of utterances. That knowledge of word meaning is different form knowledge of other things does not entail that the two types of information are isolated from one another in some special way during the moment-by-moment processing of linguistic input.

To put this in terms of the three-way distinction Borg (2004) draws (discussed above in chapters 1 and 2), we can maintain that the semantics-pragmatics distinction is grounded in the fact that knowledge of semantics is Chomsky-modular—that is, that the language faculty consists of a semantic component, a database consisting of a lexicon and rules of composition for generation of phrasal meanings.[37] The inclusion of such a database in the faculty of language would ground the distinction between semantic and pragmatic aspects of linguistic understanding. And, as I argued (chapters 1, 2), the absence of Fodor-modular semantic processing would not entail the falsity of the Chomsky-modular hypothesis regarding semantic competence.

So, there are at least two ways we can preserve the semantics-pragmatics distinction in a mentalist linguistics without appealing to the Fodor-like view of semantic processing. The first is by appeal to alternative notions of performance-oriented modularity that allow for some top-down, extra-modular effects, but still with a limited form of encapsulation. Provided that the degree of penetrability is low enough (that is, that the limited form of encapsulation is not too limited), this notion of modularity for semantic processes might allow us to draw a principled distinction between the two types of *processes*, even if the informational distinction is blurred. The second way to preserve the semantics-pragmatics distinction in a mentalist linguistics is by grounding it in facts about competence (or Chomsky-modularity) alone.

37 Two quick clarifications: As noted, this is part of what Borg claims, but she also claims that semantics is Fodor-modular, as I make clear in chapter 1's exposition. As also noted above, Pietroski's internalist "pointer" semantics is pitched at the level of competence, or Chomsky-modularity, not at the level of performance.

Suppose that we were to conclude that, from the perspective of linguistic processing, there is no principled distinction between semantics and pragmatics. Here is the picture. You hear some noises, and some reasonably encapsulated processes are responsible for phonetic and syntactic processing; so those, at least, are not inferential. But the rest is inference, aided by shortcuts and heuristics. What word are you, as speaker, uttering? What concept do you intend for me to retrieve on the way to attributing a speaker meaning to you? These questions are not decided except by way of isotropic inference guided by heuristics and contextually driven biases. No encapsulated system resolves the sense of 'book.' In general, if I end up attributing a message to you, this is the result not of semantically decoding the linguistic input in accord with my knowledge of the language used, but the result of pragmatic inference. In effect, on the picture being described here, Quine (1956: 132; cited in Rey 2020) was perhaps correct, at least when it comes to the perspective of linguistic processing: during moment-by-moment utterance interpretation, perhaps there is no principled distinction to be found between the *use* to which my knowledge of linguistic meaning is put and the use to which I put my knowledge of anything else.

Although the picture just sketched is compatible with preservation of the semantics-pragmatics distinction in a mentalist linguistics (by appeal to Chomsky-modularity), it does have certain consequences that will be of interest to many philosophers of language. A very common thought in the philosophy of language is that, when you utter a sentence, there is some message that you are in some sense responsible or accountable for. However, if the operations of the

processing system that is responsible for building and composing concepts on the way to an attribution of speaker meaning are shot through with top-down considerations about information not provided by the sentence that you have uttered, then it is unclear in what sense you, as speaker, are responsible for the messages that I attribute to you. On the radically non-modular view under consideration here, my inferences about the concept expressed by a word are subject to the influence of any given piece of information available to me, and you of course are not in a position (as speaker) to consider every piece of information available to me (as a hearer) before you produce your utterance.

We can of course appeal to considerations about what it would be rational for a hearer to infer, given the linguistic input and everything else, and cash out accountability for 'what is said' in terms of such normative notions (see Saul 2002 for more on normative versus non-normative versions of the notions of saying and implicating, discussed briefly in section 1.4.2 above). However, there is a certain attraction to the Fodor-style encapsulation hypothesis for semantics precisely because it lets us naturalize this normative notion. The Fodor-style hypothesis allows us to say that, at a minimum, the inferences of the hearer are constrained by the results of narrowly semantic decoding of the linguistic input. Resolving the reference of indexicals and working out the intended senses of polysemes require some amount of isotropic inferential work, but the linguistic input affords crucial scaffolding that constrains the space of hypotheses regarding speaker meaning, scaffolding that is itself made available, without any inferential work, by narrowly linguistic processes beyond phonetics and parsing.

However, given phenomena like polysemy, the same problem arises even if we preserve a version of the Fodor-style encapsulation hypothesis for concept retrieval. As argued, the two theories of word meaning for polysemes like 'bank$_1$' and 'book' are the chimeric (translating 'book' to a list of privileged concepts) and the pointer theories (translating 'book' to a pointer to a region in conceptual space with vague boundaries). As noted, such regular polysemy is pervasive in natural language. If there is a semantic processor that carries out the function of content resolution, and if that processor is encapsulated in the Fodor-like way, then it cannot eventuate in FOLIO given 'book' as input. The choice of FOLIO instead of STORY given recognition of 'book' cannot be made by a Fodor-style encapsulated processing system. Instead, an encapsulated stage of processing could either make available to further processing a chimeric list of the two concepts, or else a vague region in conceptual space in which the two concepts are located, along with others.

No other version of the Fodor-like hypothesis for content resolution is available regarding processing of regular polysemes. At best, there is a stage at which a handful of candidate resolutions are made available by an encapsulated system—perhaps a discrete list of privileged options (or chimera), or maybe a vague indication of a space of options (or pointer). Hence, even if we stick with the Fodor-like hypothesis for semantic processing (or some other notion of modular semantic processing), content retreats (cf. Levinson 2000) from the linguistic input, leaving inference to keep the lines of communication open.

Bibliography

Allen, Badacker, & Osterhout 2003. Morphological analysis in sentence processing: an ERP study. *Languages and cognitive processes* 18: 405-430.

Allott, forthcoming. On inference, encapsulation, and utterance interpretation. Forthcoming in *Inquiry.* [Draft as of 30 June 2021.]

Antony 2003. Rabbit-pots and supernovas: on the relevance of psychological data to linguistic theory. In *Epistemology of language*, Barber (ed.)

Apresjan 1974. Regular polysemy. *Linguistics* 14(2): 5-32.

Arnold, Eisenband, Brown-Schmidt, & Trueswell 2012. The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition* 76(1): 13-26.

Asher 2011. *Lexical meaning in context*. Cambridge.

Austin 1961. *Philosophical papers*. Oxford.

Austin 1962. *How to do things with words.* Oxford.

Azuma & van Orden 1997. Why safe is better than fast: the relatedness of a word's meanings affects lexical decision times. J. Memory Lang. 36: 484-504.

Bach 1994. Semantic slack: what is said and more. In *Foundations of speech act theory*, Tsohatzidis (ed.). Routledge.

Bach & Harnish 1979. *Linguistic communication and speech acts*. 1979. MIT Press.

Balcetis & Dunning 2010. Wishful seeing: More desired objects are seen as closer. *Psychological science* 21: 147-52.

Bar-Hillel 1971. Out of the wastebasket of pragmatics. *Linguistic inquiry* 2(3): 401-407.

Barrett et al. 1988. Event-related potentials and the matching of familiar and unfamiliar faces. *Neuropsychologia* 26: 105-117.

Barrett & Rugg 1989. Event-related potentials and the semantic matching of faces. *Neuropsychologia* 27: 913-922.

Barrett & Rugg 1990. Event-related potentials and the semantic matching of pictures. *Brain & Cognition* 14: 201-212.

Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier 1999. ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *Journal of cognitive neuroscience* 11: 235-260.

Binder & Desai 2011. The neurobiology of semantic memory. *Trends in cognitive science* 15: 527-536.

Bloomfield 1933. *Language*. University of Chicago Press.

Borg 2004. *Minimal semantics*. Oxford.

Bredart & Modolo 1988. Moses strikes again: focalization effect on a semantic illusion. *Acta psychologica* 67: 135-144.

Brown & Hagoort 1993. The processing nature of the N400: evidence from masked priming. *Journal of cognitive neuroscience* 5: 34-44.

Bruner 1957. On perceptual readiness. *Psychological review* 64: 123-152.

Burkhardt 2007. The P600 reflects cost of new information in discourse memory. *Neuroreport* 18(17): 1851-1854.

Cappelen & Lepore 2005. *Insensitive semantics: a defense of semantic minimalism and speech act pluralism*. Wiley-Blackwell.

Carston 2002a. Linguistic meaning, communicated meaning and cognitive pragmatics. *Mind and language* 17(1, 2): 127-148.

Carston 2002b. *Thoughts and utterances: the pragmatics of explicit communication*. Blackwell.

Carston 2012. Word meaning and concept expressed. *The linguistic review* 29(4): 607-623.

Carston 2019. Ad hoc concepts, polysemy, and the lexicon. In *Relevance, pragmatics, and interpretation*. Cambridge.

Carston 2020. Polysemy: pragmatics and sense conventions. *Mind and language* 36: 108-133.

Carruthers 2003. On Fodor's problem. *Mind and language* 18(5): 502-523.

Carruthers 2006. The case for massively modular models of mind. In *Contemporary debates in cognitive science*, Stainton (ed.). Blackwell.

Chierchia & McConnel-Ginet 1990. *Meaning and grammar*.

Chomsky 1964. *Current issues in linguistic theory*. Mouton.

Chomsky 1965. *Aspects of the theory of syntax*. MIT Press.

Chomsky 1971. Recent contributions to the theory of innate ideas. In *Philosophy of language*, Searle (ed.). Oxford.

Chomsky 1977. *Essays on form and interpretation*. North Holland.

Chomsky 1986. *Knowledge of language*. Praeger.

Chomsky 1992. Language and interpretation: philosophical reflections and empirical inquiry. In *Inference, explanation and other philosophical frustrations*, Earman (ed.). University of California Press.

Chomsky 1995. Language and nature. *Mind* 104: 1-61.

Chomsky 2000. *New horizons in the study of language and mind*. Cambridge.

Clark 1996. *Using language*. Cambridge.

Copestake & Briscoe 1995. Semi-productive polysemy and sense extension. *Journal of semantics* 12: 15–67.

Cruse 2004. *Meaning and language: an introduction to semantics and pragmatics*. Oxford.

Davidson 1986. A nice derangement of epitaphs. In *Truth and interpretation:*

perspectives on the philosophy of Donald Davidson. Blackwell.

Drożdżowicz 2015. Speakers' intuitive judgments about meaning: the voice of performance view. *Review of philosophy and psychology* 9(1): 1-19.

Duffy, Morris & Rayner 1988. Lexical ambiguity and fixation times in reading. *Journal of memory and language* 27: 429-46.

Duhem 1906/1964. *The aim and structure of physical theory*. Atheneum.

Eddington & Tokowicz 2015. How meaning similarity influences ambiguous word processing: the current state of the literature. *Psychonomic bulletin and review* 22: 13-27.

Elbourne 2011. *Meaning: a slim guide to semantics*. Oxford.

Embick, Hackl, Schaeffer 2001. A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive brain research* 10: 345–348.

Erickson & Mattson 1981. From words to meaning: a semantic illusion. *Journal of verbal learning and verbal behavior* 20: 540-552.

Falkum 2011. *The semantics and pragmatics of polysemy: A relevance-theoretic account*. Doctoral Dissertation. University College London.

Falkum & Vicente 2015. Polysemy: current perspectives and approaches. *Lingua* 157: 1-16.

Federmeier 2007. Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44: 491-505.

Fiorentino & Poeppel 2004 Decomposition of compound words: An MEG measure of early access to constituents. In R. Alterman & D. Kirsch (eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society*. Erlbaum.

Firestone & Scholl 2016. Cognition does not affect perception: evaluating the evidence for "top-down" effects. *Behavioral and brain sciences* 39: doi:10.1017/S0140525X15000965.

Fletcher, Happe, Frith U., Baker, Dolan, Frackowiak, Frith C.D. 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition* 57: 109-128.

Fodor 1983. *Modularity of mind: an essay on faculty psychology*. MIT Press.

Fodor 1986. The modularity of mind. In *Meaning and cognitive structure*, Pylyshyn & Demopoulos (eds.). Ablex.

Fodor 1987. Modules, frames, fridgeons, sleeping dogs, and the music of spheres. In *The robot's dilemma,* Pylyshyn (ed.). Ablex.

Fodor 1989. Schiffer's dark night of the soul: a review of *Remnants of meaning*. *Philosophy and phenomenological research* 50: 409-423.

Fodor 1998. *Concepts: where cognitive science went wrong*. Clarendon.

Fodor 2000. *The mind doesn't work that way*. MIT Press.

Fodor & Lepore 1998. The emptiness of the lexicon: reflections on James Pustejovky's *The generative lexicon*. *Linguistic theory* 29: 269-288.

Fodor & Lepore 1999/2002. Why compositionality won't go away: reflections on

Horwich's 'deflationary' theory. RUCCS technical paper. Reprinted in *Meaning and representation*, Borg (ed.). Blackwell.

Forster 1979. Levels of processing and the structure of the language processor. In *Sentence processing: psycholinguistic essays presented to Merrill Garrett*, Cooper & Walker (eds.) Erlbaum.

Foraker & Murphy 2012. Polysemy in sentence comprehension: effects of meaning dominance. *Journal of memory and language* 67(4): 407-425.

Frazier 1987. Theories of sentence processing. In *Modularity in knowledge representation and natural-language understanding*, Garfield (ed.). MIT.

Frazier & Rayner 1990. Taking on semantic commitments: processing multiple meanings vs. multiple senses. *Journal of memory and language* 29(2): 181-200.

Frisson 2009. Semantic underspecification in language processing. *Language and linguistics compass* 3(1): 111-127.

Gaston 2020. *The role of syntactic prediction in auditory word recognition*. Doctoral Dissertation. University of Maryland.

Gibbs 1984. Literal meaning and psychological theory. *Cognitive science* 8: 275-304.

Green, McCourt, Lau and Williams 2020. Processing adjunct control: Evidence on the use of structural information and prediction in reference resolution. *Glossa: a journal of general linguistics* 5(1): 112. 1-33.

Grice 1957. Meaning. *The Philosophical Review* 66(3): 377-388.

Grice 1961. The causal theory of perception. *Proceedings of the Aristotelian Society* (supplementary volume) 35: 121-52.

Grice 1968. Utterer's meaning, sentence meaning and word meaning. *Foundations of language* 4: 225-242.

Grice 1975. Logic and conversation. In *Syntax and semantics: volume 3, speech acts*, Cole & Morgan (eds.). Academic Press.

Grossi 2006. Relatedness proportion effects on masked associative priming: an ERP study. *Psychophysiology* 43: 21-30

Grossman, Rhee & Moore 2005. Sentence processing in frontotemporal dementia. *Cortex* 41: 764-777.

Hagoort 2008. The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical transactions of the royal society B: biological sciences*. 363: 1055-1069.

Harris 2020. Semantics without semantic content. *Mind & language*.

Harris, Fogal & Moss 2017. Speech acts: the contemporary theoretical landscape. In *New work on speech acts*, Harris, Fogal & Moss (eds.). Oxford.

Hauk & Pulvermuller 2004. Effects of word length and frequency on the human event-related potential. *Clinical neurophysiology* 115: 1090-1103

Hauk, Davis, Ford, Pulvermuller & Marslen-Wilson 2006. The time course of

visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage* 30: 1383-1400.

Heim and Kratzer 1998. *Semantics in generative grammar*. Blackwell.

Hodges et al. 1992. Semantic dementia: progressive fluent aphasia with temporal lobe atrophy. *Brain* 115: 1783-1806.

Holcomb 1993. Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology* 30: 47-61.

Holcomb & McPherson 1994. Event-related brain potentials reflect semantic priming in an object decision task. *Brain and cognition* 24: 259-276

Holcomb & Neville 1990. Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Language and cognitive processes* 5: 281-312.

Jackendoff 2002. *Foundations of language: brain, meaning, grammar, evolution*. Oxford.

Kaplan 1989. Demonstratives: an essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In *Themes from Kaplan*, Almog, Perry, Wettstein (eds.). Oxford.

Klepousniotou et al. 2008. Making sense of word senses: the comprehension of polysemy depends on sense overlap. *Journal of experimental psychology: learning, memory, cognition* 34(6): 1534-1543.

Kalénine, Peyrin, Pichat, Segebarth, Bonthoux, and Baciu 2009. The sensory-motor specificity of taxonomic and thematic conceptual relations: a behavioral and fMRI study. *NeuroImage* 44: 1152-1162.

Katz 1972. *Semantic theory*. Harper & Row.

Katz & Postal 1991. Realism versus conceptualism in linguistics. *Linguistics and philosophy* 14(5): 515-554.

Kempson 2001. Pragmatics: Language and communication. In *Handbook of linguistics*, Aronoff & Rees-Miller (eds.). Blackwell.

Kiefer 2002. The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive brain research* 13: 27-39.

Kintsch 1988. The role of knowledge in discourse comprehension: A construction–integration model. *Psychological review*, 95, 163–182.

Klein & Murphy 2001. The representation of polysemous words. *Journal of memory and language*. 45(2): 259-282

Klepousniotou & Baum 2007. Disambiguating the ambiguity advantage effect in word recognition: an advantage for polysemous but not homonymous words. *Journal of neurolinguistics* 20(1): 1-24.

Kosslyn 1994. *Image and brain*. MIT Press.

Kripke 1977. Speaker's reference and semantic reference. *Midwest studies in philosophy* 2: 255-276.

Kutas 1993. In the company of other words: electrophysiological evidence for

single-word and sentence context effects. *Language and cognitive processes* 8: 533-572

Kutas & Federmeier 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology* 62: 621-647.

Kutas & Hilyard 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207: 203-205.

Langacker 1991. *Concept, image, and symbol*. Mouton.

Lau, Phillips & Poeppel 2008. A cortical network for semantics: (de)constructing the N400. *Nature reviews: neuroscience* 9: 920-933.

Levinson 2000. *Presumptive meanings: the theory of generalized conversational implicature*. MIT Press.

Lewis 1970. General semantics. *Synthese* 22: 18-67.

Lewis 1979. Scorekeeping in a language game. *Journal of philosophical logic* 8(3): 339-359.

Lin & Murphy 2001. Thematic relations in adults' concepts. *Journal of experimental psychology* 130: 3-28.

Lucas 1999. Context effects in lexical access: a meta-analysis. *Memory & cognition* 27(3) 385-398.

MacDonald, Pearlmutter & Seidenberg 1994. The lexical nature of syntactic ambiguity resolution. *Psychological review* 101: 676-703.

MacGregor, Bouwsema & Klepousniotou 2015. Sustained meaning activation for polysemous but not homonymous words: evidence from EEG. *Neuropsychologia* 68: 126-138.

Matchin & Hickok 2020. The cortical organization of syntax. *Cerebral cortex* 30: 1481-1498.

Mazoyer et al. 1993. The cortical representation of speech. *Journal of cognitive neuroscience* 5: 467-479.

McGurk & McDonald 1976. Hearing lips and seeing voices. *Nature* 264: 746-748.

McRae, Spivey-Knowlton & Tanenhaus 1998. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of memory and language* 38: 283-312.

Montague 1973. The proper treatment of quantification in ordinary English. In *Approaches to natural language*, Moravcsik & Suppes (eds.). Reidel.

Neale 1990. *Descriptions*. MIT Press.

Neale 2004. This, that, and the other. In *Descriptions and beyond*, Bezuidenhout & Reimer (eds.). Oxford.

Neale 2005. Pragmatism and binding. In *Semantics versus pragmatics*, Szabó (ed.). Oxford.

Nieuwland & van Berkum 2006. When peanuts fall in love: N400 evidence for

the power of discourse. *Journal of cognitive neuroscience* 18(7): 1098-1111.

Nunberg 1979. The non-uniqueness of semantic solutions: polysemy. *Linguistic and philosophy* 3: 143-184.

Ogilvie & Carruthers 2016. Opening up vision: the case against encapsulation. *Review of philosophy and psychology* 7(4): 721-742.

Reder & Kusbit 1991. Locus of the Moses illusion: imperfect encoding, retrieval, or match? *Journal of memory and language* 30: 385-406.

Park & Reder 2004. Moses illusion. In *Cognitive illusions*, Pohl (ed.). Psychology Press.

Pallier, Devauchelle & Dehaene 2011. Cortical representation of the constituent structure of sentences. *Proceedings of the national academy of sciences*. 108: 2522-2527.

Pettit 2002. Why knowledge is unnecessary for understanding language. *Mind* 111(443): 519-550.

Pietroski 2006. Character before content. In *Content and modality: themes from the philosophy of Robert Stalnaker*, Thomson & Byrne (eds.). Oxford.

Pietroski 2008. Minimalist meaning, internalist interpretation. *Biolinguistics* 4: 317-341.

Pietroski 2010. Concepts, meanings, and truth: *Mind and language* 25: 247-278.

Pietroski 2017. Semantic internalism. In *The Cambridge companion to Chomsky*, McGilvray (ed.). Cambridge.

Pietroski 2018. *Conjoining meanings: semantics without truth values.* Oxford.

Pustejovsky 1995. *The generative lexicon*. MIT Press.

Putnam 1975. The Meaning of 'Meaning'. In *Language, mind and knowledge*, Gunderson (ed.). University of Minnesota Press.

Pylkkänen 2015. Composition of complex meaning: interdisciplinary perspectives on the left anterior temporal lobe. In: Hickok G, Small S, editors. *Neurobiology of language*, Hickok & Small (eds.). Academic Press.

Pylkkännen & Marantz 2003. Tracking the time course of word recognition with MEG. *Trends in cognitive sciences* 7: 187–189.

Pylkkännen & Murphy 2006. The representation of polysemy: MEG evidence. *Journal of cognitive neuroscience* 18(1): 1-13.

Pylyshyn 1999. Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and brain sciences* 22(3): 341-65.

Quine 1956. Carnap and Logical Truth. In his *Ways of paradox and other essays*. Harvard University Press.

Quine 1960. *Word and object.* MIT Press.

Quine 1969. Epistemology naturalized. In his *Ontological relativity and other essays*. Columbia University Press.

Rayner et al. 1983. The interaction of syntax and semantics during sentence

processing: eye movements in the analysis of semantically biased sentences. *Journal of verbal learning and verbal behavior* 22: 358-374.

Recanati 2004. *Literal meaning*. Cambridge.

Rey 2020. *Representation of language: philosophical issues in a Chomskyan linguistics*. Oxford.

Rodd, Gaskell & Marslen-Wilson 2002. Making sense of semantic ambiguity: semantic competition in lexical access. *Journal of memory and language* 46(2): 245-266.

Rogalsky & Hickok 2008. Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral cortex* 19: 786-796.

Rugg 1985. The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology* 22: 642-647.

Rugg & Nagy 1987. Lexical contribution to nonword-repetition effects: evidence from event-related potentials. *Memory and cognition* 15: 473-481.

Ruhl 1989. *On monosemy: a study in linguistic semantics*. SUNY Press.

Saul 2002. 'Speaker meaning, what is said, and what is implicated.' *Noús* 36 (2) 228-248.

Schiffer 2003. *The things we mean*. Oxford.

Schumacher 2011. The hepatitis called…: electrophysiological evidence for enriched composition. In *Experimental pragmatics/semantics,* Meibauer & Steinbach (eds.). John Benjamins.

Schumacher 2014. Content and context in incremental processing: "the ham sandwich" revisited. *Philosophical studies* 168: 151-165.

Searle 1975. Indirect speech acts. In *Syntax and semantics, vol. 3: speech acts*, Cole & Morgan (eds.). Academic Press.

Searle 1980. The background of meaning. In *Speech act theory and pragmatics*, Searle, Keifer & Bierwisch (eds.). Reidel.

Sereno et al. 1998. Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport* 9: 2195-2200.

Sikos, Duffield, and Kim 2016. Grammatical predictions reveal influences of semantic attraction in online sentence comprehension: evidence from speeded forced-choice sentence continuations. *Language, cognition and neuroscience* 1-19.

Spalek 2015. Spanish change of state verbs in composition with atypical theme arguments: clarifying the meaning shifts. *Lingua* 157: 36-53.

Sperber 1996 *Explaining culture: a naturalistic approach*. Blackwell.

Sperber & Wilson 1986/1995. *Relevance: communication and cognition*. Blackwell.

Srinivasan & Rabagliati 2015. How concepts and conventions structure the lexicon: cross-linguistic evidence from polysemy. *Lingua* 157: 124-152.

Stainton 2009. The role of psychology in the philosophy of language. In

Routledge companion to the philosophy of language, Russell & Graff Fara (eds.). Routledge.

Strawson 1950. On referring. *Mind* 59(235): 320-344.

Swinney 1979. Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of verbal learning and verbal behavior* 18: 645-660.

Thompson, Bonakdarpour, Fix, Blumenfeld, Parrish, Gitelman & Mesulam 2007. Neural correlates of verb argument structure processing. *Journal of cognitive neuroscience* 19: 1753-1767.

Travis 1985. On what is strictly speaking true. *Canadian journal of philosophy* 15.

Trueswell, Tanenhaus & Garnsey 1994. Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *Journal of memory and language* 33: 285-318.

Trueswell and Tanenhaus 1994. *Approaches to studying world-situated language use: bridging the language-as-product and language-as-action traditions*. MIT Press.

Tyler and Marslen-Wilson 1977. The on-line effects of semantic context on syntactic processing. *Journal of verbal learning and verbal behavior* 16: 683-692.

Van Berkum 2008. Understanding sentences in context: what brain waves can tell us. *Current directions in psychological science* 17(6): 376-380.

Van Berkum et al. 1999. Semantic integration in sentences and discourse: Evidence from the N400. *Journal of cognitive neuroscience* 11: 657-671.

Van Gompel & Pickering 2017. Syntactic parsing. In *The Oxford handbook of psycholinguistics*, Gaskell (ed.). Oxford.

Van Oostendorp & de Mul 1990. Moses beats Adam: a semantic relatedness effect on a semantic illusion. *Acta psychologica*, 74: 35-46.

Van Oosterndorp & Kok 1990. Failing to notice errors in sentences. *Languages and cognitive processes* 5: 105-113.

Van Petten & Kutas 1990. Interactions between sentence context and word frequency in event-related brain potentials. *Memory and cognition* 18: 380-393.

Vicente 2018. Polysemy and word meaning: an account of lexical meaning for different kinds of content words. *Philosophical studies* 175: 947-968.

Wilson & Carston 2007. A unitary approach to lexical pragmatics: relevance, inference and ad hoc concepts. In *Pragmatics*, Noel Burton-Roberts (ed.). Palgrave.

Wilson, DeMarco, Henry, Gesierich, Babiak, Mandelli, Miller, Gorno-Tempini 2014. What role does the anterior temporal lobe play in sentence-level processing? Neural correlates of syntactic processing in semantic variant primary progressive aphasia. *Journal of cognitive neuroscience* 26: 970-985.

Warren 1970. Perceptual restoration of missing speech sounds. *Science* 167: 392-393.

Wittgenstein 1953/2009. *Philosophical investigations*. 4th edition. Hacker & Schulte (eds. and trans.). Oxford.

Zwicky & Saddock 1975. Ambiguity tests and how to fail them. In: *Syntax and semantics, vol. 4*, Kimball (ed.). Academic Press.