The Language of Thought

A New Philosophical Direction

The Language of Thought

A New Philosophical Direction

Susan Schneider

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Stone by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Schneider, Susan, 1968–
The language of thought : a new philosophical direction / Susan Schneider.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-01557-8 (hardcover : alk. paper)
1. Philosophy of mind. 2. Cognitive science—Philosophy. 3. Thought and thinking—Philosophy. 4. Fodor, Jerry A. I. Title.
BD418.3.S36 2011
153.4—dc22

2010040700

10 9 8 7 6 5 4 3 2 1

To Rob, Jo, Denise, and Ally

Contents

Preface

This project started when certain of the language of thought program's central philosophical commitments struck me as ill conceived. It might have ended after several lengthy arguments with Jerry Fodor, but I am more stubborn than he is.

The idea that the mind is computational pervades contemporary cognitive science and philosophy of mind. Within cognitive science, it has become something like a research paradigm. And over the years, I've been very happy with that research paradigm—thrilled, actually. Who would deny that the last thirty or so years have witnessed an amazing beginning for cognitive science? But I must confess that computationalism's philosophical credentials always struck me as weaker than the science behind it. For what is it to say that the mind is computational? We cannot merely assume that if the *brain* is computational, the *mind* is as well. There are substance dualists who accept the former while repudiating the latter, after all. No, we need to reflect on whether the mind is computational even on the assumption that computationalism about the *brain* is promising. Here, philosophers have ventured two sorts of computational approaches to the mind: one that is based on a connectionist, or neural network, approach, and one—the language of thought (LOT)

approach—that takes thinking to consist in the algorithmic manipulation of mental symbols.

Now, I thought to write a book-length exposé of the flaws in connectionist approaches to higher cognitive function, but someone already had (Marcus 2001). And in any case, it struck me that, philosophically speaking, connectionism is actually far better off than LOT, for its leading proponents are at least bona fide computationalists. Fodor, in contrast, is not. So I decided to sit down and ponder the scope and limits of the LOT approach, to determine if it is even a well-conceived computational approach to begin with. In this book, I do not intend to rule out non-computationalist options (e.g., biological naturalism, substance dualism): I trust many readers have arrived at views on this matter; they pick up this book because they find computationalism about the mind to be prima facie attractive. Yet even to those who sympathize with the computational approach, LOT seems to be in deep philosophical trouble: in the last several years, numerous cracks have emerged in its conceptual foundations. Its theory of meaning conflicts with its theory of computation; its theory of concepts is too emaciated—too nonpsychological—to be a satisfactory theory of concepts; Fodor's recent books on LOT actually argue that the cognitive mind is noncomputational; and even LOT's conceptual cornerstone—the very notion of a symbol—is poorly understood.

So here, I grapple with these problems, and at the end of the philosophical day, I believe you will find that the LOT I arrive at is quite different from the orthodox philosophical LOT. For the new LOT seeks integration with cognitive and computational neuroscience—indeed, LOT's naturalism requires it. And I repudiate Fodorian pessimism about the capacity of cognitive science to explain cognition. Further, in my hands LOT becomes

a *pragmatist* theory: I argue that LOT couldn't have been otherwise, and that even the mainstream, Fodorian LOT made hidden appeals to pragmatism, while officially embarking on a massive attack on it, quite ironically. Relatedly, I advance a pragmatist version of conceptual atomism: *pragmatic atomism*.

I imagine that you will care about all this if you've signed on to the LOT program. And if you are vehemently opposed to LOT, you may want to know whether the LOT you are opposed to is really one that requires all the philosophical wares commonly associated with it, which you've come to know and hate. I am claiming that LOT is different than you think.

But before I launch into all this, allow me to give credit where credit is due. First and foremost, I would like to thank Jerry Fodor for his many thought-provoking ideas, and for numerous philosophical discussions. I'm afraid he will disagree with much of this book, but I hope my reworking of LOT inspires fruitful lines of inquiry. I am also grateful to the National Endowment for the Humanities for their financial support, to Philip Laughlin at MIT Press for his efficient editing and helpful advice, to Melanie Mallon and Katherine Almeida at MIT Press for their thorough copyediting, and to the audiences at various departments who hosted me at their colloquia in which chapters of this book were presented (the University of Maryland, Washington University at St. Louis, the University of Pennsylvania, Lehigh University, and the University of Cincinnati).

This book drew from several earlier papers of mine: "The Nature of Symbols in the Language of Thought," *Mind and Language* (Winter 2009): 523–553; "LOT, CTM and the Elephant in the Room," *Synthese* (Winter 2009): 235–250; "Fodor's Critique of the Classical Computational Theory of Mind" (with Kirk Ludwig), *Mind and Language* 23 (2008): 123–143; "Direct Reference,

Psychological Explanation, and Frege Cases," *Mind and Language* 20, no. 4 (September 2005): 223–447; "Conceptual Atomism Rethought," *Behavioral and Brain Sciences*, 33 , pp 224–225; and "Yes, It Does: A Diatribe on Jerry Fodor's Mind Doesn't Work That Way," *Psyche* 13, no. 1 (Spring 2007): 1–15. I would like to thank the editors and reviewers at these journals for their useful suggestions.

I am especially grateful to Mark Bickhard, Gary Hatfield, John Heil, Michael Huemer, and Gerald Vision. Not only did they give insightful feedback on parts of the manuscript, but they provided valuable practical advice and words of encouragement as well. I am also very grateful to the following people for their helpful comments on certain chapters: Murat Aydede, David Braun, Adam Croom, Matt Katz, Jonathan Cohen, Frances Egan, Michael Huemer, Brian McLaughlin, Carlos Montemayor, Jesse Prinz, Philip Robbins, Andreas Scarlatini, Murray Shanahan, Whit Schonbein, Bradley Rives, Jacob Beck, and Gualtiero Piccinini. The work of many of these people has played a significant role in the development of this book. Kirk Ludwig was also key to this project, to say the least, as he coauthored one of its chapters. I've enjoyed working with Ludwig, and indeed, all of these people, immensely. Needless to say, despite help from such a stellar crowd, I am sure errors have inevitably crept in, and that these are all due to me.

Last but most significantly, I am grateful to my family. I am especially indebted to my mother-in-law, Jo Marchisotto, and sister-in-law, Denise Marchisotto, who watched my young one while parts of the book were being written, and to both my husband and daughter, Rob and Alessandra Marchisotto, who tolerated an all-too-often distracted writer in their midst.

1 Introduction

Minds, whatever these are, are the bearers of mental states. And it is the primary ambition of philosophy of mind to figure out the nature of minds and their states. No matter what one thinks of the language of thought program, it is clear that it offers an influential theory of the nature of thoughts and the minds that have them. With respect to minds, the program says that they are symbol-manipulating devices of an ultrasophisticated sort. With respect to mental states, these are said to be mental symbols—ways in which we conceive of the world—strung together by operations of an inner grammar, the behavior of which is to be detailed by a completed cognitive science.

And this brings me to the very reason that I sat down to write this book. Despite the language of thought program's enormous influence, I wasn't sure that if you thought any of this through—if you hadn't already—you would say that you *really* know what it means. Hence, the three central puzzles of this book, which I shall now describe.

Problem One. Consider what the language of thought (LOT) approach says about the nature of mind, at least within philosophical circles. Jerry Fodor, the main philosophical advocate of LOT and the related computational theory of mind (CTM),

claims that while LOT is correct, the cognitive mind is likely noncomputational (2000, 2008). This is perplexing, to say the least, because LOT is supposed to be a computational theory, and CTM quite obviously is supposed to be such as well (Fodor 1975).[1] Further, as I'll illustrate, insofar as Fodor even entertains the idea that cognition is computational, he employs a view in which cognitive processing is entirely sequential, and in which the biological underpinnings of cognition are ignored. Herein, I dismantle Fodor's case against computationalism, and provide LOT with a superior account of the computational character of the cognitive mind, at least in rough outline.

Problem Two. What about mental states? Of course, LOT and CTM say that mental states are *symbolic*; but what is a symbol? Strangely, LOT, a position that claims that thinking is *symbol* processing, has never clarified what a symbol is (Anderson 2007, Ch 1; Marcus 2001, 147). Indeed, the issue has been "largely neglected" (Pessin 1995, 33). Yet symbols are the conceptual cornerstone of the LOT program. They are supposed to capture our ways of conceiving the world, figure as kinds in explanations of thought and behavior, and enable LOT to integrate thought into the world that science investigates. Indeed, Fodor has voiced the worry that without a coherent theory of symbols ". . . the whole [LOT] project collapses" (Pessin 1995, 33). I agree. In this book I single out a conception of a symbol for LOT—it is the only notion of a symbol suitable to play the important philosophical

1. CTM holds that the mind is computational, with thinking being the algorithmic manipulation of semantically interpretable symbols in LOT. LOT and CTM are close kin—so much so that CTM generally consists in a philosophical commitment to LOT, together with an added semantic dimension that many build into LOT in any case. For this reason, following Fodor, I'll refer to the *philosophical* program surrounding both as simply the "LOT program," or simply "LOT."

and scientific roles that symbols are summoned to play. Yet once symbol natures are understood, elements of the current philosophical LOT program must be discarded. And LOT becomes a pragmatist theory.

Problem Three. Central to any account of the nature of mentality is a story about the representational nature of thought. So how is it that LOT's mental states come to represent, or be about, entities in the world? Here, many advocates of LOT appeal to a theory of meaning or mental content that is (to a first approximation) referential: e.g., the content or meaning of both the mental symbols #Cicero# and #Tully# is just the man, Cicero, despite the fact that the symbols differ in their cognitive significance.[2] (This kind of content has been called *broad content*). But here's the rub: as I'll explain, this approach to content conflicts with the view that thinking is symbolic, at least insofar as a leading, neo-Russellian theory of belief ascription is employed; and this is the view that mainstream LOT currently favors (Aydede and Aydede 1998; Aydede and Robbins 2001; Fodor 1994; Schneider 2005). For the standard LOT faces *Frege cases*: counterexamples to intentional generalizations arising from intentional laws that are sensitive to broad contents, rather than being sensitive to the particular ways the individual conceives of the referent. Frege cases undermine LOT's ability to explain thought and behavior. Further, Frege Cases suggest that LOT's account of mental states is deeply mistaken: mental states may very well be symbolic and computational, but it is unclear whether, as such, they can also

2. Although LOT's position on content is referential in the case of proper names, indexicals, and demonstratives, note that the content of a predicate is a property, rather than an extension at a world. The literature on LOT tends to ignore this subtlety (and I'll follow). Note that I will designate LOT expressions by enclosing the relevant expression with the symbol "#" (e.g., #dog#).

have the sort of semantic features that the standard LOT claims that they have.

I must confess that these three problems have been grating on me for some time, so permit me a rant: I have an ambivalent relationship toward LOT and CTM. They are immensely problematic, at least in their philosophical incarnation, being both divorced from the rest of cognitive science and plagued by serious problems. And I am quite interested in connectionist and dynamical systems approaches to the brain. Yet something seems unfair about LOT's philosophical predicament: in cognitive science proper, the symbol-processing view is alive and well. The notion of a symbol requires clarification, admittedly, but overall, there are sensible symbolic approaches in cognitive science, including, for instance, Gary Marcus's recent survey of why connectionism, if it is to explain cognition, must employ symbolic resources (2001). Even an influential computational neuroscience textbook respects the need for symbolic resources in models of higher cognitive function (O'Reilly and Munakata 2000). And the symbol-processing approach is the leading view of the format of thought in information-processing psychology; in this domain, LOT is thriving.

But as you've surely noticed, the philosophical LOT troubles me. For one thing, it turns away from the symbol processing tradition in cognitive science, as I've indicated. For Fodor, as brilliant as he is, is at heart not a computationalist. His latest books include several chapters arguing that LOT's *central system*—LOT's expression for the system responsible for higher cognitive function—will likely defy computational explanation (2000, 2008).[3] Astonishingly, he implores cognitive science to stop research on

3. Fodor also advances this position at the end of an earlier book, *The Modularity of Mind* (1983).

the central system. For another thing, the standard LOT wages an all-out war with concept pragmatism. In the literature on concepts, LOT famously opposes "pragmatist" accounts of the nature of thought, where by "pragmatist views" Fodor means claims that one's abilities (e.g., one's recognitional, classificatory, or inferential capacities) determine the nature of concepts (Fodor 2004, 34). In fact, Fodor proclaims that pragmatism is the "defining catastrophe of analytic philosophy of language and philosophy of mind in the last half of the twentieth century" (2003, 73-74). And he declares in his *LOT 2* that LOT's favored theory of concepts and, indeed, LOT itself, represents an important alternative to pragmatism (2008, 12).

After careful consideration of Fodor's positions on these matters, I will argue that Fodor situates LOT on the wrong side of both of these battles; indeed, I will develop a LOT that is firmly rooted in both pragmatism and computationalism. Back in 1975, Fodor noted in his *The Language of Thought* that characterizing the language of thought "is a good part of what a theory of mind needs to do," and this classic book was a brilliant exposition and defense of the symbolicist position in cognitive science (1975, 33). But a good deal of work still needs to be done. So here is what I propose to do: my reflections on Problem One will give rise to a LOT that is squarely computationalist, being integrated with current findings in cognitive science. And my solution to Problem Two will reveal that LOT cannot oppose pragmatism, for a symbol is individuated by what it does, that is, by its psychological role: the role it plays in one's mental life, including the role it plays in recognition, classification, and inference.[4] This leaves LOT with an account of symbols—LOT's neo-Fregean

4. In particular, symbols shall be individuated by their role in computation.

"modes of presentation"—that is pragmatist in an important sense. Further, this result extends to LOT's theory of concepts (*conceptual atomism*) as well, generating a superior version of the view. And concerning Problem Three, this new understanding of symbols, together with argumentation that I suspect is even palatable to the current LOT, provides a solution to the problem of Frege cases.

It is largely due to Fodor's insights that philosophers of mind have come to appreciate LOT. But if you ask me, LOT has been straying from its rightful path. Lost afield, it cries out for a philosophical overhaul.

Some Background

But I am skipping ahead a bit, presupposing background knowledge of the terrain. I should now like to provide a brief survey of the LOT approach for those who are relatively uninitiated. Then—for my connectionist friends and other critics of the LOT program—I should like to make clear why it is worth bothering to devise solutions to these three problems, refurbishing LOT, rather than just throwing our hands up in the face of them and embracing one of *their* views instead. For I suspect that if you are a critic of LOT, then this is your reaction to our three problems. I shall then trace the dialectical path through which the chapters proceed.

According to the LOT program, conceptual thinking occurs in an internal languagelike representational medium. However, this internal language is not equivalent to one's spoken language(s). Instead, LOT is the format in which the mind represents concepts. The LOT hypothesis holds that the mind has numerous mental "words" (called *symbols*) that combine into

mental sentences according to the grammatical principles of the language. When one thinks, one is engaged in the algorithmic processing of strings of these mental symbols.[5] The LOT program and the connectionist program are often viewed as competing theories of the format, or representational medium, of thought.[6]

As you may have surmised, the idea that there is a language of thought is commonly associated with the work of Jerry Fodor, who defended this hypothesis in an influential book, *The Language of Thought* (1975), and who has continued to do so in the context of a steady and influential stream of books and articles.[7] The philosophical literature on LOT focuses on the LOT program as it is developed by Fodor, in which the idea that we think in an inner symbolic language is developed in tandem with a constellation of related issues concerning meaning, modularity, concepts, CTM, and more. Fodor's writings on these topics are commonly part of the graduate and undergraduate canon in philosophy of mind. I mention this because readers in other fields of cognitive science may not realize that within philosophy, Fodor's program *is* for the most part the philosophical face of LOT and CTM. This book is a philosophical treatment of the language of thought approach, so quite naturally, it is a close

5. By "algorithm" I mean an effective, step-by-step procedure that manipulates strings of symbols and generates a result within finitely many steps.

6. But this latter issue is actually more subtle than this, as I explain shortly. For very readable overviews of connectionism, see Churchland (1996), which focuses on philosophical issues, and Hawkins (2005). The latter author is a scientist who provides a fairly up-to-date, broad-ranging discussion of the connectionist-based approach to intelligence.

7. Some would say that the LOT position dates back to Plato, who obviously was not a computationalist, but who held that each of us has innate concepts of universals, or forms, that we recall, at least to a degree.

treatment of, and reaction to, Fodor's ideas. This being said, the reader should bear in mind that those theorists who ascribe to LOT but reject the views that happen to be under fire in a given chapter (e.g., anti-computationalism) are not the targets of my criticisms.

According to Fodor, LOT was inspired by the ideas of Alan Turing, who defined computation in terms of the formal manipulation of uninterpreted symbols according to algorithms (Turing 1950; Fodor 1994). In Turing's "Computing Machinery and Intelligence," he introduced the idea that symbol-processing devices can think, a view that many in cognitive science are sympathetic to, yet which has also been the focus of great controversy (e.g., Searle 1980; Dreyfus 1972; Turing 1950). Indeed, the symbol-processing view of cognition was very much in the air during the time when Fodor's *Language of Thought* was published (1975). Two years before the publication of *The Language of Thought*, Gilbert Harman published his *Thought*, in which he argued that mental states "have structure, in the way that sentences have structure. . . . Mental states are part of a system of representation that can be called a language of thought" (1973, 65).[8] And three years before *The Language of Thought* came out, Allen Newell and Herbert Simon suggested that psychological states could be understood in terms of an internal architecture that was like a digital computer (Newell and Simon 1972). Human psychological processes were said to consist in a system of discrete inner states (symbols), which are manipulated by a central processing unit (CPU). Sensory states serve as inputs to the system, providing the "data" for processing according to the

8. Harman's book does not wed LOT to pessimism and anti-pragmatism, and I am sympathetic to it.

rules, and motor operations serve as outputs. This view, called *classicism*, was the paradigm in the fields of artificial intelligence, computer science, and information-processing psychology until the 1980s, when the competing connectionist view also gained currency. LOT, as a species of classicism, grew out of this general trend in information-processing psychology to see the mind as a symbol-processing device. The classicist tradition stressed an analogy between cognition and digital computers while downplaying the relevance of neuroscience to understanding cognition. Even today, the symbol-processing approach is at the heart of information-processing psychology and philosophy of mind, being one of two leading computational theories of the nature of thought (the other being connectionism).

Now, let us ask: Why believe in the language of thought? The most important rationale for LOT derives from the following observation: any empirically adequate theory of mind must hold that cognitive operations are sensitive to the constituent structure of complex sentencelike representations (Fodor 1975; Fodor and Pylyshyn 1995; Marcus 2001). This observation has been regarded as strong evidence for a LOT architecture. Consider the sentence "The cappuccino in Italy is better than in China." Despite never hearing this sentence before, you are capable of understanding it. Thought is *productive*: in principle, you can entertain and produce an infinite number of distinct representations. How can you do this? Our brains have a limited storage capacity, so we can't possibly possess a mental phrase book in which the meaning of each sentence is encoded. The key is that the thoughts are built out of familiar constituents and combined according to rules. It is the *combinatorial* nature of thought that allows us to understand and produce these sentences on the basis of our antecedent knowledge of the grammar and atomic

constituents (e.g., *China*, *Italy*). This allows for the construction of potentially infinitely many thoughts given a finite stock of primitive expressions (Chomsky 1975; Fodor 1975, 31; Fodor and Pylyshyn 1988, 116; Fodor 1985, 1987).

Relatedly, consider the phenomenon of systematicity. A representational system is *systematic* when the ability of the system to produce (or entertain) certain representations is intrinsically related to the ability to produce (or entertain) other representations (Fodor and Pylyshyn 1995, 120). Conceptual thought seems to be systematic; e.g., one doesn't find normal adult speakers who can produce "Mary loves John" without also being able to produce "John loves Mary." How can this fact be explained? Intuitively, "Mary loves John" is systematically related to "John loves Mary" because they have a common constituent structure. Once one knows how to generate a particular sentence out of primitive expressions, one can also generate many others that have the same primitives (Fodor 1987; Fodor and Pylyshyn 1988; Fodor and McLaughlin 1990).

Systematicity and productivity are commonly regarded as providing significant motivation for LOT. Whether any connectionist models can explain these important features of thought is currently very controversial (see, e.g., Fodor and Pylyshyn 1988; Fodor and McLaughlin 1990; Elman 1998; van Gelder 1990; Marcus 2001; Smolensky 1988, 1995). Connectionist models are networks of simple parallel computing elements with each element carrying a numerical activation value, which the network computes given the values of neighboring elements, or units, in the network, employing a formula. In very broad strokes, critics claim that a holistic pattern of activation doesn't seem to have the needed internal structure to account for these features of thought (Marcus 2001; Fodor and Pylyshyn 1988). Critics have

argued that, at best, certain connectionist models would model how symbol structures are implemented in the brain; they cannot really represent genuine alternatives to the LOT picture, however (Fodor and Pylyshyn 1988). There is currently a lively debate between this "implementationalist" position and radical connectionism, a position that advances connectionism as a genuine alternative to the language of thought hypothesis.

Now let us turn to a more detailed discussion of LOT's fundamental claims. We've noted that LOT holds, first and foremost, that thinking is the algorithmic manipulation of mental symbols. This view, when fleshed out more fully, is generally taken to involve the following three claims.

(1) *Cognitive processes consist in causal sequences of tokenings of internal representations in the brain.*

Rational thought is said to be a matter of the causal sequencing of tokens—patterns of matter and energy—of representations that are realized in the brain. Rational thought is thereby describable as a physical process, and further, as we shall see below, as both a computational and a semantic process as well.

In addition:

(2) *These internal representations have a combinatorial syntax and semantics, and further, the symbol manipulations preserve the semantic properties of the thoughts* (Fodor 1975; Fodor and Pylyshyn 1988).

Although technically, the LOT hypothesis does not require that symbols have a semantics, in practice, that symbols have a semantics has, in effect, become part of many elaborations of LOT (but see Stich, 1994). This being said, claim (2) has three components:

(2a) *Combinatorial syntax*.

As noted, complex representations in the language of thought (e.g., #take the cat outside#) are built out of atomic symbols (e.g., #cat#), together with the grammar of the language of thought.

(2b) *Combinatorial semantics*.

The meaning or content of a LOT sentence is a function of the meanings of the atomic symbols, together with their grammar.

(2c) *Thinking, as a species of symbol manipulation, preserves the semantic properties of the thoughts involved* (Fodor 1975; Fodor and Pylyshyn 1988).

To better grasp (2c), consider the mental processing of an instance of *modus ponens*. The internal processing is purely syntactic; nonetheless, it respects semantic constraints. Given true premises, the application of the rule will result in further truths. The rules are truth preserving. John Haugeland employs the following motto to capture this phenomenon:

Formalist Motto: If you take care of the syntax of a representational system, the semantics will take care of itself (Haugeland 1989, 106).

And finally:

(3) *Mental operations on internal representations are causally sensitive to the syntactic structure of the symbol* (Fodor and Pylyshyn 1988).

Computational operations work on any symbol/symbol string satisfying a certain structural description, transforming it into another symbol/symbol string that satisfies another structural description. For example, consider an operation in which the system recognizes any operation of the form (P&Q) and transforms it into a symbol of the form (P). Further, the underlying

physical structures onto which the symbol structures are mapped are the very properties that cause the system to behave in the way it does (see Fodor and Pylyshyn 1988, Macdonald 1995, ch. 1; Marcus 2001, ch. 4; Smolensky 1988, 1995).[9]

Claims (1)–(3) are the primary tenets of the LOT position. Further, they underlie a view that is closely related to LOT, the aforementioned computational theory of mind (or "CTM").

CTM: Thinking is a computational process involving the manipulation of semantically interpretable strings of symbols, which are processed according to algorithms (Newell and Simon 1976; Fodor 1994; Pinker 1999; Rey 1997).

Steven Pinker nicely captures the gist of the manner in which (1)–(3) give rise to CTM:

> Arrangements of matter . . . have both representational and causal properties, that is, . . . [they] simultaneously carry information about something and take part in a chain of physical events. Those events make up a computation, because the machinery was crafted so that if the interpretation of the symbols that trigger the machine is a true statement, then the interpretation of the symbols created by the machine is also a true statement. The Computational Theory of Mind is the hypothesis that intelligence is computation in this sense. (1999, 76)

This statement aptly connects the CTM hypothesis to the classic question, "How can rational thought be grounded in the brain?" According to LOT and CTM, rational thought is a matter of the causal sequencing of symbol tokens that are realized in the brain

9. It turns out that this feature of classical systems—that the constituents of mental representations are causally efficacious in computations—plays a significant role in the debate between LOT and connectionism, for in contrast to symbolic systems, connectionist systems do not operate on mental representations in a manner that is sensitive to their form.

(thesis 1). These symbols, which are ultimately just patterns of matter and energy, have both representational (thesis 2b) and causal properties (thesis 3). Further, the semantics mirrors the syntax (thesis 2c). This leaves us with the following picture of the nature of rational thought: thinking is a process of symbol manipulation in which the symbols have an appropriate syntax and semantics (roughly, natural interpretations in which the symbols systematically map to states in the world).

Advocates of LOT and CTM mine this account of the nature of rational thought in their attempt to solve an important puzzle about intentional phenomena. By *intentional phenomena* I mean a thought's "aboutness" or "directedness"—that it represents the world as being a certain way. Thought has long been suspected of being somehow categorically distinct from the physical world, being outside the realm that science investigates. For how is it that thoughts (e.g., the belief that the espresso is aromatic, the desire to drink Merlot), which, as we now know, arise from states of the brain, can be about, or directed at, entities in the world? In essence, advocates of LOT and CTM approach this question in a *naturalistic* way, trying to ground intentionality in the scientific world order. Now, I've already noted that symbols have a computational nature. As such, they are part of the scientific domain. But the proponent of LOT has a naturalistic story about the aboutness, or intentionality, of symbols as well: Symbols refer to, or pick out, entities in the world in virtue of their standing in a certain causal or nomic relationship to entities in the world. Simply put, the symbols are "locked onto" properties or individuals in virtue of standing in a certain nomic or causal relationship specified by a theory of meaning or mental content. So the intentionality of a thought is a matter of a nomic

or causal, and ultimately physical, relationship between mental symbols and entities in the world.

This naturally brings me to the matter of LOT's standard line on semantics. I have just mentioned, somewhat vaguely, that the LOT program posits a "locking" relationship between symbols and referents. As it turns out, LOT's semantic side is multifaceted. Bearing in mind that Problem Three concerns the relationship between symbols and psychological generalizations sensitive to semantic properties, the reader may benefit from more detail.

Proponents of LOT are generally *externalists* about mental content, holding that thought content is not solely determined by one's internal states; instead, content depends upon entities in the world as well. Further, advocates of LOT often have a specific sort of externalist mental content in mind, *broad content*, which, as observed, is basically referential.[10] Why would one hold this unintuitive view? One reason is that individuals having different types of inner states—in the context of LOT, this amounts to different types of symbolic states—can nonetheless have thoughts with the same content. Content individuation becomes a species of the metaphysics of object and property individuation. To be sure, these are big ticket items in metaphysics, but every theory of meaning relies on a manner of individuating such entities in any case. Worries that content will be as idiosyncratic as thought threatens to be drop out of the picture. Different thinkers, and indeed, different kinds of minds, can, at least in principle, have thoughts that refer to the same entities.

10. Here I'm following Fodor's usage of "broad content" (Fodor 1994, 7). There is another usage in which it is taken as being synonymous with *wide content*.

I've mentioned that LOT's semantic story is multifaceted. In addition to appealing to broad content, the standard LOT adopts a related position called *neo-Russellianism*. According to neo-Russellianism, the proposition expressed by the sentence "Cicero is Tully" is an entity that consists in the relation of identity, the man, Cicero, and the man, Tully. Further, the sentence "Tully is Tully" expresses the same proposition. On this view, individuals are *literally* constituents of neo-Russellian propositions. As Bertrand Russell commented to Gottlob Frege, "Mont Blanc itself is a component part of what is actually asserted in the proposition 'Mont Blanc is more than 4000 meters high'" (Frege 1980, 169). Crucially, neo-Russellians hold that "believes" expresses a binary relation between agents and propositions. They therefore hold the surprising view that anyone who believes that Tully is Tully also believes that Cicero is Tully.[11] I explore neo-Russellianism in chapter 8; for now, observe that a LOT that is wedded to neo-Russellianism adopts the following claim about psychological explanation:

(PE) Sentences in one's language of thought that differ only in containing distinct primitive co-referring symbols (e.g.,

11. This view has also been called the *naïve theory*, *Russellianism* and *Millian-Russellianism*. By *neo-Russellianism*, I am of course not referring to Russell's other, descriptivist, account of names. Neo-Russellianism has been defended by (*inter alia*) David Braun, Keith Donnellan, David Kaplan, Ruth Marcus, John Perry, Mark Richard (in one incarnation, at least), Bertrand Russell, Nathan Salmon, Scott Soames, and Michael Thau. In addition to providing an account of proper names, neo-Russellians typically extend their account to other expression types. To keep matters simple, I focus on the case of proper names. For a helpful overview of neo-Russellianism and other theories of attitude ascription, see Richard (1990).

#Cicero#/#Tully#) are to be treated by intentional psychology as being type-identical and are thereby subsumable under all the same intentional laws (Fodor 1994).

The position sketched in these last paragraphs is often referred to as *referentialism*.[12]

Now, nothing in this book argues that referentialism, or even externalism, is correct, although I am sympathetic to these positions, at least when it comes to the content of thought. Again, the task of this book is to determine the scope and limits of the LOT program—to lay bare its key problems, to respond to them, and in the course of advancing these responses, to identify new contours to the intellectual landscape. In this vein, the third problem concerns whether referentialism is even compatible with LOT to begin with, for Frege cases suggest that LOT's neo-Russellian-inspired generalizations—that is, generalizations conforming to (PE)—face counterexamples and are thus not viable. If Frege cases cannot be solved, something is deeply wrong with LOT's current thinking about mental states.

Discarding LOT?

I have now discussed the basic elements of the LOT program, including the family of doctrines that have come to be associated with the LOT position in its philosophical incarnation: naturalism, intentionality, neo-Russellianism, broad content, and more. Bearing all this in mind, recall our earlier discussion of the three problems that LOT faces. These problems threaten the very

12. In this book I reserve the expression "referentialism" for the neo-Russellian position, not the hidden indexical theory, although the hidden indexical view also appeals to broad content.

fabric of the LOT program: its drive to explain thinking computationally in terms of the manipulation of mental symbols deriving their meaning, or "aboutness," from nomic relations to the world. Surely if the cognitive mind is not even computational, if the very notion of a symbol is empty, or if its intentional generalizations have counterexamples, LOT should be discarded.

Perhaps the weight of these problems is simply too staggering. Given the prominence of connectionism, for instance, shouldn't the verdict be that we must finally put LOT to rest? There will be some who are eager to draw this conclusion; I suspect, however, that doing so would be premature. For one thing, cognitive science is increasingly making progress on higher cognitive function, and models of higher cognition seem to be precisely the terrain in which one would expect to see validation of the symbol-processing approach, if validation is to come. As connectionists Randall O'Reilly and Yuko Munakata admit in a recent computational neuroscience textbook, the symbolic approach to higher-level cognition has a "long history of successful models," for "in symbolic models, the relative ease of chaining together sequences of operations and performing arbitrary symbol binding makes it much more straightforward to simulate higher-level cognition than in a neural network." In contrast, "neural network models of higher-level cognition are in their relative infancy" (O'Reilly and Munakata 2000, 379). And although representation in the prefrontal cortex (PFC) is still poorly understood relative to many other brain regions, as they point out, representation in the PFC appears to be combinatorial and discrete. If this turns out to be correct, it would support an appeal to symbolic models to explain higher-level cognition (perhaps implemented by connectionist networks, perhaps not). The combinatorial and discrete representations of

the PFC are distinct from the more distributed modality-specific representation of the posterior cortex; prima facie, this latter representation seems more straightforwardly amenable to traditional connectionist explanation.[13] And all this comes from the latest computational neuroscience textbook, not just from symbolicists like Fodor and Marcus.

Add to this the fact that the precise relationship between LOT and connectionism is extremely subtle. The proponent of LOT has an important rejoinder to the connectionist attempt to do without mental symbols: to the extent that the connectionist can explain the combinatorial nature of thought, then connectionist systems would, at best, merely provide models in which symbols are implemented in the cognitive mind. Such systems do not really represent genuine alternatives to the LOT picture. For the networks would ultimately be the lower-level implementations of symbolic processes. This view was briefly mentioned earlier; it is often called *implementational connectionism*.[14]

Further, our reflections shall have implications for philosophy of cognitive science more generally, bearing on the scope and limits of naturalistic approaches to the mind. For instance, any sort of theory of mind appealing to referentialism will likely face the third problem, which involves the presence of counterexamples to intentional generalizations. And although other approaches will not experience problems with the nature of mental *symbols*, similar issues with the individuation of mental states arise in the context of both narrow content and connectionism, and computationalists of any stripe face the kind of problems that motivated

13. For a discussion of the distinct processing in the PFC and posterior cortex, see O'Reilly and Munakata (2000, 214–219).

14. For discussion, see Fodor and Pylyshyn (1988), Pinker and Prince (1988), Marcus (2001).

Fodor to conclude that the cognitive mind is likely noncomputational (Churchland 2005; Fodor and LePore 1992; Fodor 2000). These issues will not simply disappear on their own.

Hence, the insight that symbolicism is still central, even for those who appreciate connectionism, is my point of departure. Fortunately, I suspect that the problems before us can be solved, and in what follows, I provide a synopsis of how the book shall endeavor to do so.

Overview of the Chapters

Making progress on important theoretical questions, such as that of the nature of mind, involves discarding inadequate theories as well as generating new, more plausible ones. This being said, little reason remains for upholding a LOT framework while also denying that conceptual thought is itself computational. Fodor may be right: the central system may not be computational, but then thought simply isn't symbolic. For this reason, chapters 2 and 3 aim to solve the first problem, setting aside Fodor's pessimistic concerns and developing a genuinely computational approach to the central system.

Fodor's pessimism is motivated by two problems that are commonly known as the relevance and globality problems. Chapter 2 focuses on the relevance problem: the problem of if and how humans determine what is relevant in a computational manner. Fodor suggests that there is an absence of viable computational approaches, indicating that the cognitive mind is likely noncomputational. I contend that a more fruitful way to proceed is to assume that the presence of a relevance problem in humans is not terribly different from relevance problems confronting other computational systems. In the case of human cognition,

however, the "solution" is a matter of empirical investigation of the brain mechanisms underlying human searches. I then sketch the beginnings of a solution to the relevance problem that is based on the global workspace (GW) theory, a theory of consciousness that extends to higher cognitive function more generally, and that is well received in psychology, cognitive neuroscience, cognitive robotics, and philosophy.[15]

Chapter 2 then develops a positive account of the central system, outlining a computational theory that is based on the GW approach. I provide only an outline because the research I discuss is just now under development. Still, a distinctive element of this approach is that it frames a LOT that embraces work in cognitive and computational neuroscience to sharpen its understanding of the central system—in fact, as mentioned, I urge that LOT *must* pay close attention to neuroscience if it is to be a bona fide naturalistic theory. Further, I argue that the popular doctrine of multiple realizability should not discourage interest in neuroscience on the part of the proponent of LOT.

The defense of LOT from pessimistic worries continues into chapter 3, where Kirk Ludwig and I respond to Fodor's globality problem. *Global properties* are features of a sentence in the language of thought that depend on how the sentence interacts with a larger plan (i.e., a set of LOT sentences). Fodor believes that the fact that thinking is sensitive to such properties indicates that thought is noncomputational. In response, Ludwig and I argue that not only is Fodor's version of the globality problem self-defeating but other construals of the problem are also highly problematic.

15. The GW theory is also called the *global neuronal workspace theory*. Researchers in neuroscience often use this expression instead.

Chapters 4, 5, and 6 tackle the second problem, developing a theory of symbols and employing it to reshape LOT. I propose that symbol natures are a matter of the role they play in computation, and more specifically, they are determined by the symbol's *total computational role*—the role the symbol plays in the algorithm that is distinctive to the central system. I will call this position on symbol natures the *algorithmic view*. The algorithmic view is not new: in the past, both Fodor and Stich have appealed to it (Fodor 1994; Stich 1983). However, because both of their discussions were extremely brief, neither philosopher offered arguments for the position nor defended it from objections. And, as will become evident from my discussion of his objections to my proposal, Fodor came to repudiate it.[16] My case for the algorithmic view is twofold: first, chapter 4 identifies the central and commonly agreed upon philosophical functions that symbols are supposed to play and determines whether any competing conceptions do in fact fill these nonnegotiable roles. The answer is: none do.

Then, more positively, chapter 5 provides three arguments for the algorithmic view. Readers familiar with debates over functionalism about mental states are likely already concerned about the notorious problem of *publicity* that attaches to individuating mental states in terms of their total functional or computational roles. Indeed, both Fodor and Prinz have separately responded to the arguments I offer in chapter 5, employing an argument that I call the *publicity argument*, which contends that because symbols are determined by the role they play in one's entire cognitive economy, different individuals will not have symbols of the same type. For people differ in the memories and cognitive

16. Stich himself is no longer sympathetic to LOT.

abilities they possess—indeed, even the same person may do so at different times.[17] Generalizations that are sensitive to symbols will not then be "public": different individuals, or even the same individual at distinct times, will not satisfy the same generalizations (for a similar criticism, see Aydede 2000).

In chapter 6, I offer an extensive reply to this objection, as well as responding to related objections, and in doing so, I also illustrate how my theory of symbols reshapes LOT. Here, I develop LOT's approach to neo-Fregean modes of presentation, which LOT views as nonsemantic and, in particular, symbolic. And in chapter 7 I summon the algorithmic conception of symbols to devise a theory of concepts. *Pragmatic atomism* is a version of conceptual atomism that draws from both concept pragmatism and referentialism, believe it or not. I argue that pragmatic atomism is far superior to mainstream conceptual atomism, for it can satisfy more of the desiderata that many believe a theory of concepts should satisfy. This is due to the fact that pragmatic atomism introduces a much-needed psychological element to conceptual atomism.

Finally, chapter 8 turns to the third problem, the problem of Frege cases. When individuals lack certain knowledge that is relevant to the success of their behaviors, they can fail to behave as LOT's neo-Russellian-based intentional laws predict, for again, such laws are sensitive to broad contents and are insensitive to the particular ways in which the referents are represented. Critics suggest that Frege cases illustrate that psychological explanation must be sensitive to one's ways of conceiving things (Aydede

17. Fodor and Prinz separately offered this objection in personal correspondence and discussion. This issue is closely connected to debates concerning functional role theories of concept and content individuation (see, e.g., Fodor and LePore 1992, Fodor 2004).

and Aydede 1998; Aydede and Robbins 2001). In this chapter, I attempt to solve this problem, and in doing so, I put the algorithmic conception of symbols to work to further refine the LOT program's account of the causation of thought and behavior. I also provide some background on the relation between broad content and theories of belief ascription, such as neo-Russellianism and the hidden indexical theory; it is crucial that readers appreciate fully why Frege cases arise in the first place, and why solving them is important to many philosophical advocates of LOT.

I would be delighted if the discussion of Frege cases, together with the chapters on symbol individuation, proved useful to philosophers of language interested in the nature of propositional attitude ascription. The attitude ascription literature, as sophisticated as it is, tends to work against the backdrop of a rather impoverished conception of cognition in which vague discussions of "guises" and "tokens in belief boxes" are commonplace. As philosophers of language know, these expressions are fudge words that urgently need fleshing out. It is my hope that this book spells out these important ideas in the context of one influential theory of mind.

Reconfiguring the Language of Thought Approach

This, then, is my game plan. Now assume for a moment that my budget of solutions *works*. One thing should be clear already: this reconfigured LOT aspires to explain the computational nature of the central system instead of holding the suicidal position that computationalism stops at the modules. Other results are equally significant. For example, LOT was developed in the absence of a theory of symbols, despite the ironic fact that LOT's

key contention is that cognition just *is* the processing of symbols. On the other hand, the theory of symbols that I defend *reshapes* LOT. For if this book is correct, then LOT finally has a concrete theory of modes of presentation (or "MOPs"), as recall, LOT's MOPs are just symbols. Moreover, once symbols are individuated, computational theories can, at least in principle, better determine how proprietary entities (e.g., activation patterns) relate to symbolic processes. Further, the algorithmic view generates an improved version of conceptual atomism.

And now, I would like to offer a disclaimer. This book is obviously not intended to be a full treatment of current answers to the question, "If and how is the brain plausibly computational?" For instance, it does not treat the connectionist or dynamical approaches to cognition. But one must approach massive questions piecemeal: LOT is one influential approach to answering the question, so let's see if it is even a good approach. So what I offer you herein is an assessment of the scope and limits of the LOT program, including considerations for rethinking certain key issues that the program currently addresses. Thus, in lieu of a firm endorsement of LOT and CTM, I venture that, assuming the problems laid out herein can be tackled, the program offers an important and prima facie plausible proposal concerning the nature of conceptual thought. Of course, I happen to suspect that the solutions work and that this is indeed the present state of things; barring that, I am happy to write an exposé of the problems that LOT and CTM face. At the very least, it will inspire a new appreciation of the problem space. That's progress too.