The Nature of Symbols in the Language of Thought

SUSAN SCHNEIDER

Abstract: The core of the language of thought program is the claim that thinking is the manipulation of symbols according to rules. Yet LOT has said little about symbol natures, and existing accounts are highly controversial. This is a major flaw at the heart of the LOT program: LOT requires an account of symbol natures to naturalize intentionality, to determine whether the brain even engages in symbol manipulations, and to understand how symbols relate to lower-level neurocomputational states. This paper provides the much-needed theory of symbols, and in doing so, alters the LOT program in significant respects.

According to the Classical Computational Theory of the Mind (simply called, 'CTM'), expressions in the language of thought will play a key role in a completed scientific theory of mind. For thinking is supposed to be a computational process involving the manipulation of symbols in the language of thought (or 'LOT'). However, as central as LOT symbols are to the CTM program, it has long been observed that the proponents of LOT have yet to provide a plausible account of the nature of primitive expressions in the language; that is, it is claimed that LOT lacks a defensible condition on the type individuation of such expressions.[1] It is fair to say that without a plausible theory of the primitives, it is unclear how the mind is supposed to be computational, in a classical sense: for after all, in absence of a theory of primitive symbols, the computational theory will not be able to determine what type a given token belongs to. And further, without a grasp of symbol natures, it would be unclear how patterns of neural activity could be, at some higher level of abstraction, accurately described as being symbol manipulations, for what is it

Thanks very much to Jerry Fodor, whose work inspired this project, for his helpful comments on various drafts of this essay and for numerous heated debates on this topic. Thanks also to the following audiences for comments on this paper: the departments of philosophy at the University of Cinncinatti, Temple University, and the University of Maryland, the cognitive science department at Lehigh University, the audience at the ESPP/SPP joint meeting in Barcelona, and the audience at the Central APA in Chicago. Special thanks to an anonymous reviewer at this journal, Gordon Bearn, Mark Bickhard, John Bickle, Peter Carruthers, Margaret Cuonzo, Cris Gauker, Kirk Ludwig, Tom Polger, Jesse Prinz, Gualtierro Puccinnini, Gary Hatfield, Georges Rey, Bill Robinson and Gerald Vision for helpful discussions and/or written comments. And thanks very much to Murat Aydede for his helpful reply at the APA meeting.

Address for correspondence: 423 Claudia Cohen Hall, University of Pennsylvania, 249 South 36th Street, Philadelphia, PA 19104-6304, USA.
Email: susansdr@gmail.com

[1] This problem has been discussed by Aydede, 2000; Devitt, 1991; Fodor, 1994; Ludwig and Schneider, 2008; Pessin, 1995; Prinz, 2002; Schneider 2009; 2005; Stich, 1983, 1993. As Murat Aydede notes, 'This... question has in fact been around, constantly popping up here and there and haunting people working in the field, for more than fifteen years, mostly thanks to Stich and Fodor's early ''methodological solipsism''' (2000).

Mind & Language, Vol. 24, No. 5 November 2009, pp. 523–553.

that is being manipulated? It would thereby seem difficult for cognitive scientists working on symbolic processing to determine if LOT is, in fact, correct. Further, some cognitive scientists hold that connectionist networks implement symbolic computations; others claim that the brain is part symbolic and part connectionist, being a hybrid device.[2] Perhaps so. But how do we know that connectionist networks relate to symbols in any of these ways if we do not even know what symbols are? Different views may simply be talking past each other, employing distinct conceptions of symbols to begin with.

To make matters worse, without a theory of symbols, symbols cannot do the important philosophical work that they have traditionally been summoned to do for the LOT program. Philosophers and other cognitive scientists have taken computational entities like symbols (in the language of thought) and activation patterns (across connectionist networks) to provide a notion of a mode of presentation (or 'MOP') that is purely narrow, or 'in the head', being determined by the intrinsic properties of the system. Such entities have been employed to account for the distinct ways that we think of referents or states of affairs. In addition, because proponents of CTM turn to LOT to naturalize intentionality, without a theory of symbols, they will be unable to provide a story about the computational basis of the internal vehicle of thought. And, over and above these worries, there is the familiar ontological concern that one cannot meaningfully posit the existence of an entity without providing a principle for its individuation.

Clearly, this situation does not bode well for LOT and CTM. Indeed, things are so bad that Jerry Fodor, the founding father of LOT, has voiced the worry that '. . . the whole [LOT] project collapses unless some coherent account of syntactic type identity can be provided.'[3] Yet, ironically, the problem of type individuation has been characterized as being 'largely neglected'.[4] I have recently detailed problems with each of the existing proposals for symbol natures (Schneider, 2009b). Herein, I further explore this neglected topic—this time venturing a positive theory of the nature of the symbols that is designed to overcome existing objections. My claim will be that CTM requires a theory that types tokens by sameness and difference of total computational role, where the total computational role of a symbol is understood as the role it plays in the algorithms of a completed cognitive science.[5]

---

2 Fodor and Pylyshyn, 1988; Wermter and Sun, 2000.

3 Quoted in Pessin, 1995, p. 33 (from personal communication with Fodor). It should be added that semantics is also impacted by the issue of symbol types because those who are interested in LOT frequently say that meaning is determined by some sort of external relation between symbols and contents. Since symbols are the internal vehicle that the meanings lock onto, they are significant to such theories.

4 Pessin, 1995, p. 33. Pessin attributes the neglect to the attention being directed toward the semantics of the symbols (p. 33). I agree.

5 In the past, a similar position has been occupied by both Jerry Fodor (1994) and Stephen Stich (1983), although it is fair to say that the view was not adequately developed. Stich had appealed to individuation of syntax by computational role in his well-known defense of syntactic eliminitivism (1983). Unfortunately, very little elaboration and defense of this manner of individuation was provided. And Jerry Fodor, in his first appendix to *The Elm and*

My discussion will consist of three parts. In part one, I provide three arguments for the individuation of symbols by their total computational roles. The first of these arguments claims that Classicism *requires* that primitive symbols be typed in this manner. The second argument contends that without this manner of symbol individuation, there will be cognitive processes that fail to supervene on symbols, together with the rules (i.e. rules of composition and other algorithms). This situation is very problematic for CTM, as CTM holds that cognitive processing just *is* the processing of mental symbols, according to rules. The third argument says that cognitive science needs a natural kind that is typed by total computational role. Otherwise, either cognitive science will be incomplete, or its generalizations will have counterexamples. If any of these arguments are correct then my theory of symbols is non-negotiable for the LOT theory. If LOT is to appeal to symbols at all—which of course it must—then, like it or not, symbols must be individuated by their total computational roles. Then, in part two, I defend my account from a criticism, offered by both Jerry Fodor and Jesse Prinz, who have responded to the arguments of this paper with the charge that because different individuals will not have symbols of the same type, they will not be subsumed under the same psychological generalizations. As a result, generalizations sensitive to symbols will not be 'public'; that is, different individuals will not satisfy the same psychological generalizations.[6]

I offer a threefold reply to Fodor and Prinz: first, I disentangle psychological explanation, which mainly proceeds by functional decomposition, from the too strong requirement that psychological explanation requires that systems have at least some of the same LOT symbols. Functional decomposition does *not* require that two systems have any of the same LOT symbols in their database. Second, I point out that explanation that is sensitive to the broad content of the mental state plays a crucial role in cognitive science. But as with functional decomposition, explanation covering different systems can occur without the systems having the same internal states. Third, I explain that generalizations involving LOT states do not, by and large, quantify over particular symbol types; rather, they only quantify over symbols in general. So different individuals frequently do fall under the same generalizations in virtue of their LOT states. And finally, I observe that the only situations in which LOT symbols are subsumed under laws with respect to their particular symbol types (as opposed to being subsumed in virtue of simply having some LOT symbol or other) involves explanation which, by its very nature, appeals to the detailed workings of a particular system. And in such situations, it is inappropriate to call for symbol types that are shared across distinct sorts of systems. Thus, as far as I

the Expert, (1994), had tried to individuate symbols by the role they played in computations of a Turing machine (pp. 108-9).

6 Fodor (correspondence and discussion). Prinz (personal correspondence and discussion). See also Aydede, 2000 and Schneider, 2009b. 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.) Here, critics charge that psychological explanation will not be 'public'; that is, it will not feature explanations that cover different systems.

can tell, Fodor and Prinz's charge of 'publicity violation' is entirely benign: the publicity failure pertains to the states, but it doesn't extend to the actual cognitive explanations themselves.

Part three briefly explores the notion of computational role that is involved in the three arguments and suggests an individuation condition that meets the demands of the arguments. Finally, part four closes by outlining the ways in which this conception of symbols alters the very face of the LOT program.

1. The Proposal

Before laying out the proposal, I should quickly explain a few terms that are often employed in discussions of the nature of LOT symbols. Such symbols are frequently called 'items in LOT syntax'. However, the term, 'LOT syntax' is a bit misleading because it is not used in a way that is synonymous with the familiar use of 'syntax' as 'grammar'. Instead, 'LOT syntax' consists in both (i), the rules of composition of the language of thought, and (ii), the class of its expressions in the language, both primitive and complex.

As the present task is to individuate the primitive vocabulary items, a more precise term that encompasses only the vocabulary items would be helpful. So henceforth, I will mainly set 'syntax' aside, and speak of 'symbols' in the language of thought. But it should be noted that 'symbol' is sometimes used in a way that this project will ultimately not endorse. Symbols are sometimes taken as being entities that are both computational *and* semantic in the following strong sense: they are said to have their contents essentially. But how to individuate the symbols is precisely what is at stake in the present discussion. It is up for grabs whether the primitive 'syntactic' items should be individuated by contents, or merely by computational features. Indeed, most proposals for individuating the primitive vocabulary items do not, in fact, take content as individuative.[7] Externalists about semantics, in particular, would, by and large, not be interested in such a view, as the items in LOT are paradigmatically narrow.

Luckily, a non-semantic use of 'symbol' is not entirely inappropriate; for one also hears talk of 'uninterpreted symbols' and by and large, symbols are paradigmatically regarded as being narrow, and computational, having interpretations assigned to them. This is because meanings are, on the standard view, regarded as being irrelevant to the formal workings of a computational system.[8] Herein, I will use 'symbol' in this purely computational sense: on my view, (as we shall see) symbols are individuated by computational features alone; they also have semantic features, but such features are not essential. Thus, it is open, on my view, that type identical

[7] Works which canvass and criticize the various proposals are Aydede, 2000; Pessin, 1995; Prinz, 2002; Schneider, 2009b.

[8] For an articulation of this common view see Haugeland, 1985, p. 50. For a dissenting view see Wilson, 1997.

symbols can be assigned distinct interpretations. (This might be, for example, the computationalist's interpretation of the famous Twin Earth case.)[9] I will also sometimes speak of primitive 'words' (i.e. symbols), in the language of thought.

This being clarified, let us ask now pose the question at hand: What are the conditions for a given token to be of a particular symbol type? Or, put slightly differently, what are the conditions for determining syntactic types? I will focus on answering these questions for simple term types—the non-semantic correlates of concepts—rather than complex expressions (including entire sentences) in LOT. Further, I will assume that (for the language of thought, at least) complex expressions can be constructed from combinatorial operations on the simples. I begin by laying out what I believe are three conclusive arguments for typing symbols by their total computational roles.

The Three Arguments

The Argument from Classicism

My first argument draws from basic tenets of Classicism. According to Classicism, a mechanism computes when it produces outputs (or 'tokens') given certain input tokens, according to an algorithm. On the classical view, it is stipulated to be the nature of a token that:

(T1) Within a given program, any token can be substituted by another token of the same type in that operation without changing the computation.

This point is often taken to be trivial. Indeed, this is a basic feature of the classical view that is found in even elementary discussions of Classicism. For instance, John Haugeland introduces the concept of types by analogy with chess pieces. He underscores that pieces of the same type must function in the same way within the program; interchanging them makes no computational difference. 'Formal tokens are freely interchangeable if and only if they are the same type. Thus it doesn't make any difference which white pawn goes on which white-pawn square; but switching a pawn with a rook or a white pawn with a black one could make a lot of difference' (Haugeland, 1989, p. 52).

But evidently, those working on theories of LOT symbol individuation have overlooked the significance of these rather platitudinous discussions; for they provide a decisive reason to accept a condition that types symbols by their total computational role. For (T1) yields a condition that says that it is necessary for that for two LOT tokens to be type identical that they have same total computational roles. For (T1) says, put a bit differently:

9 On my view, the molecular duplicates have the very same internal, computational states. This accounts for the sense in which they are psychologically similar, but the symbols can map to different broad contents when the relevant features of the environment differ (Schneider, 2005). I believe that Fodor disagrees: at least in nomologically necessary worlds, symbols must map to the same broad contents (Fodor, 1994). It is a rich issue whether syntax and semantics 'stack up' for the externalist. However, it is an issue that I must leave for another time.

(T2) Any token of the same type will generate the same (proximal) output and internal states of the machine, given that the same internal states of the machine are activated.

And this yields the following necessary condition on two symbol tokens being type identical:

(T3) ∀x ∀y(x and y are type-identical LOT tokens, then x and y will have the same total computational role).

Further, according to Classicism, substituting token of type S1 for token of type S2 in a string results in a different computational process, which, assuming the process runs it course, will produce at least one different output or internal state (Haugeland, 1989, 52). This yields the following:

(T4) ∀x ∀y(if x and y are type-distinct LOT tokens, then, x and y have distinct total computational roles).

By further logical operations we the following sufficient condition on type identity:

(T5) ∀x ∀y(if x and y have the same total computational role then, x and y are type-identical).

So, by stating basic facts about how Classicism views the nature of symbols, we have made explicit the commitments that Classicism has to their type individuation. Those offering various theories on how primitive symbols in the language of thought should be typed have failed to note that there really is no room for negotiation. For we have located both a necessary and a sufficient condition on being a LOT symbol that seem to be required by Classicism. Being typed by sameness and difference of total computational role is just what it is to be a classical 'token'!

A simple and clear elaboration of these points is found in Haugeland's well-known discussion of the classical theory of computation, in *Artificial Intelligence, the Very Idea* (1985). In the second chapter, he lays out the notion of a classical computation, which he identifies as a sort of formal system in which there is a 'token manipulation game'. Here, he explains the notion of a token by analogy with chess pieces:

> Ultimately, the rules are what determines the types of the tokens. The rules specify what moves would be legal in what positions. If interchanging two particular tokens wouldn't make any difference to which moves were legal in any position, then it couldn't make any difference to the game at all; and those tokens would be of the same type. To see the significance of this, consider chess again. In some fancy chess sets, every piece is unique; each white pawn, for instance, is a little figurine, slightly different from the others. Why then are they all the same type? Because, in any position whatsoever, if you interchanged any two of them, exactly the same moves would be legal. That is, each of them

> contributes to an overall position in exactly the same way, namely, in the way that pawns do. And that's what makes them all pawns (1989, p. 52).

The upshot: Classicism requires a symbol to be typed by the role it plays in the program, that is, by sameness and difference of total computational role. How will any other notion of a symbol work with Classicism? Individuation by total computational role seems to be axiomatic.[10]

Now let us turn to the Supervenience Argument.

The Supervenience Argument

Consider the following reductio argument: Assume that within a given system, two primitive symbol tokens, *a* and *b*, are of symbol type 1, but *a* has the following causal relation that *b* does not have:

(T) Causing a tokening of LOT sentence S1 under circumstances C.

Let us ask: how can we explain the causal difference between *a* and *b*? (T) is obviously a phenomenon that is of interest to CTM as it involves cognitive, and moreover, symbolic processing. Now, according to CTM, computations are entirely determined by:

(i) the type identity of the primitive symbols in the language of thought;
(ii) the grammar of the language of thought—that is, the rules of composition that yield compound symbols in the language, including sentences;
(iii) the 'rules' or algorithms that are supposed to describe cognitive processes and predict behaviors.[11]

As noted, the grammatical properties of the language, together with the totality of symbol types, are commonly referred to as the 'LOT syntax'. Now, we've just supposed that there is a computational difference between tokens *a* and *b*, although they are type identical symbols. But notice that this difference in causal powers clearly does not translate into a difference in either (i) or (ii), for we've assumed that the tokens are type identical symbols and our assumed difference in causal powers between *a* and *b* should not impact the grammatical properties of the language.

---

[10] Perhaps Haugeland's chess example has not really been ignored by proponents of LOT, (as I suggest at infra, p. 000) but has been rejected because it leads to holism. But rejecting it for this reason would be question begging: for how can this rejection be compatible with a continued appeal to LOT? For the present point is that individuation by total computational role seems axiomatic. Proponents of LOT who would like to reject symbol holism had better start worrying about this issue.

[11] Two notes concerning (iii): First, I shall take 'algorithm' to be a particular line of code or short sequence of lines of code, which itself may be part of a larger program. Computer scientists also take algorithms as being equivalent to a complete program; philosophers of mind often mean just a line in the larger program. Herein I follow philosophers of mind. Nothing hangs on this. Second, strictly speaking, the grammatical rules mentioned in (ii) fall into the broader category of (iii) as well. But it is useful to single out the grammatical rules for the purpose of discussion.

Turning to (iii), would this difference in causal powers perhaps be a matter of a difference in the rules? Since there is a difference in causal powers between *a* and *b*, and since the causal powers in question involve mental processing, might there be some algorithm that token *a* can figure in, and which token *b* cannot? Unfortunately, the Argument from Classicism already ruled out this possibility: if *a* and *b* are type identical, they must be capable of functioning the same way in the classical program. We can also state the issue in terms of laws, rather than algorithms.[12] Since, *ex hypothesi*, *a* and *b* are type identical symbols, and according to the proponent of LOT, the computational laws are supposed to be sensitive to symbol types, any computational law which subsumes *a* will also subsume *b*. Thus, there will not be a 'special computational law' which *a* satisfies and which *b* does not, and which thus captures the causal difference between *a* and *b*. Laws, like classical algorithms, are insensitive to causal differences between tokens of the same type. The upshot: it seems that we have mental processing that is not determined by LOT syntax (including the grammar), together with the algorithms, contra CTM. Hence, to preserve CTM, we must regard *a* and *b* as being type identical symbols. The reductio leaves us with the following constraint on LOT symbol individuation:

> (Principle P) ∀x ∀y (It is not possible for x and y to be tokens in the same system, of the same type, and differ in their computational roles).

This constraint just says that sameness in (total) computational role is a necessary condition on the type identity of LOT symbols.[13]

This then, is the Supervenience Argument. What is the significance of this failure to supervene?[14] As Jerry Fodor notes in his *The Mind Doesn't Work That Way* (2000), if there is mental processing that does not supervene on syntax, then CTM is incomplete, for it cannot explain such mental processing. Recall that in the present context, the feature of cognition which LOT cannot explain is (T):

> (T) Causing a tokening of LOT sentence S1 under circumstances C.

---

[12] It is not clear that computational laws and algorithms really differ all that much: the algorithms provided by a completed cognitive science purport to state nomologically necessary generalizations and are good candidates for being computational laws.

[13] Unlike the Argument From Classicism, the Supervenience Argument does not establish the sufficiency of any condition that proceeds by sameness/difference of total computational role. However, the sufficiency is not really an issue, as the condition admittedly slices symbols extremely thinly.

[14] In his (2000) Fodor provides an entirely different argument to the conclusion that CTM is incomplete because there is mental processing which fails to supervene on syntax. Fodor concludes from the putative fact that LOT cannot explain certain features of cognition ('global' features) that the central systems are not computational; CTM only holds with respect to the modules. (Global properties are features that a mental sentence has which depend on how the sentence interacts with a larger plan, i.e. set of sentences, rather than the type identity of the sentence alone.) Elsewhere, Kirk Ludwig and I have disputed Fodor's argument but I agree with Fodor that this kind of supervenience failure, if it really ensued, would be devastating for CTM. (For further discussion see section three of this paper, Ludwig and Schneider, 2008; and Schneider, 2007.)

Ironically, what (T) is a case of, and which goes unexplained, is symbolic processing itself. Now, it would be unproblematic if LOT merely was incomplete in the sense that it needs to be supplemented by a non-symbolic theory of sensory states. That computational processing of the sort carried out in, e.g., early vision, is non-symbolic is both plausible and unproblematic. But it is another thing entirely when the explananda are cognitive phenomena, and in particular, why one LOT sentence causes another. Relatedly, if LOT is incomplete as an account of symbolic processing, then it would be unclear how intentionality is to be naturalized, as projects that attempt to naturalize intentionality that appeal to LOT clearly look to LOT for a complete account of symbolic processing. For intentional phenomena which defy explanation at the level of the computational theory would be prima facie mysterious—unnaturalizable.

Now let us turn to a third argument. As noted, this argument says that psychology needs a natural kind that is individuated by the role it plays in one's cognitive economy. This argument will speak to the case of connectionism, and to narrow mental state individuation more generally. However, it will be formulated in terms of the language of thought and Classicism.

The (Computational-Level) Frege Cases Argument

In the spirit of molecularism about content individuation, the molecularist about LOT symbols singles out certain of the symbol's computational relations as being type-individuative. The computational relations appealed to are selected in such a way as to yield symbols that are shared, from person to person. The laws the symbols figure in are there by able to subsume an extensive equivalence class of systems which, when tokening a given symbol in common conditions, behave in the same way.

I believe there is a reasonable case for the view that at least some molecularist types exist, and further, that in at least some cases, they span a large population of individuals. Here, I have in mind mental symbols for logical and mathematical expressions. In the context of discussions of the plausibility of conceptual role semantics (CRS), they are generally the parade cases. However, the problem is that the case for molecularist symbols is weak for other expression types. Molecularist theories of narrow content have failed in a related attempt to identify certain conceptual or inferential roles as being constitutive of narrow contents. In broad strokes, the problem is that there is no principled way to distinguish between those elements of conceptual or inferential role that are meaning constitutive from those which are not (Fodor and LePore, 1992). Similar issues would emerge for molecularism about symbol types, although the issues would not concern meanings; instead, the issue would be whether there could really be a select few symbol constitutive computational relations. However, in light of the counterexamples that have arisen in the narrow content literature, we have reason to be skeptical that such accounts will succeed.

To see what goes wrong, consider the following strategy to provide symbol constitutive computational relations. There might be a population of 'novices' who all know the same small amount of information about a kind. Let us employ talk of 'mental files'. Now, in the mental file for a natural or artificial kind concept, a novice may only have the most very basic facts. For example, someone may know only that 'spruce' names a type of tree. Other novice tree recognizers may have this very sparse grasp of this kind of tree as well. So the proponent of molecularism about symbols may ask: why can't a mental word for spruces be shared between the novices; i.e. those who also have this very skeletal spruce concept? A similar scenario can apply to names; consider the contrast between an expert's knowledge of Einstein and a novice who may only know that he's the person who devised relativity theory.

While such a project may initially inspire hope, what ruins the prospects for interpersonal molecularist types for symbols for names and kind terms is that insofar as there is any difference in computational role between tokens of identical symbol types, counterexamples to computational laws can ensue. (And further, as I will explain, such counterexamples cannot plausibly be taken to be tolerable exceptions.) For let us now consider the Computational-level Frege Cases Argument. This argument contends that without individuation of LOT symbols by total computational role, either there will be missed predictions or there will be 'computational-level Frege cases'. By a 'Frege case' I mean a certain sort of counterexample that arises for psychological laws that subsume states that are individuated in a manner that is too coarsely grained for the purposes of capturing important behavioral similarities. Frege cases are well-known in the literature on broad content and intentional explanation.[15] To consider a well-known Frege Case, consider Sophocles' Oedipus, who didn't realize that a woman he wanted to marry, 'Jocasta', happened to be his mother. Oedipus has two distinct MOPs, or ways of representing the same person, and he doesn't realize that they co-refer. In the literature on mental content, this situation creates problems if intentional laws are Russellian, or sensitive to broad (roughly, referential) content. For the laws are indifferent to Oedipus' distinct ways of conceiving things, and Oedipus threatens to be a counterexample to the broad generalization:

(M) Ceteris paribus, if people believe that they shouldn't marry Mother and they desire not to marry Mother, they will try to avoid marrying Mother.

Oedipus satisfies the antecedent of (M) but fails to satisfy the consequent since, in virtue of his trying to marry Jocasta, it is true, according to a broad content psychology, that he tries to marry Mother. Now, Frege cases can arise at the computational level as well. In general, Frege cases are situations in which an agent satisfies the antecedent of a psychological generalization, but fails to satisfy the consequent because the theory treats mental representations as being type identical that are, in fact, causally distinct in the way the mental representations function in the system's cognitive economy.

15 For more on intentional-level Frege cases see: Aryo, 1996; Aydede and Robbins, 2001; Fodor, 1994; Schneider, 2005; Rupert, 2008.

This being said, the strategy behind the Computational-level Frege Case argument is the following: Assume that P is some principle of individuation of LOT primitives and that P is not equivalent to individuation by total computational role. Wouldn't there be cases in which two LOT expressions are type identical according to P while differing in their total computational roles? If there aren't, then P is just equivalent to a principle that types LOT primitives in terms of sameness and difference of total computational role. But as I will argue below, if there is a case in which two LOT tokens of the same type differ in total computational role, then either there will be missed predictions or there will be Frege cases.

Here's the argument: Let 'CR' denote the causal role of a given LOT token, *a*. And let 'CR*' denote an individuation condition for the type that *a* is a token of. CR* is a condition that employs individuation by computational role, where the computational role includes every computational-level causal relation that the token enters into with other primitives except for one relation, R*. So R* is not individuative of the type that *a* is a token of. But *a* has R*. I take it that the causal relations that specify the computational role of a given token are detailed by the computational laws. So there is a given computational law, L, which specifies R*. Now, let *b* be a token that only has the causal role specified by CR*, and not CR, because *b* lacks R*. And let us suppose that like *a*, *b* is typed by CR*. Then, either: (i), both *a* and *b* will *not* be subsumable in L. Or, (ii), they will both be subsumable in L. In the case of (i) the theory will have a missed prediction: it will miss that *a* has a causal relation that is specified by L. (ii) Now consider the second scenario, in which they will both be subsumable in L. In this case *b* does not have the causal relation detailed by L. So we wouldn't expect it to behave in accordance with L. Hence, *b* will be a counterexample to L.[16]

Further, this sort of counterexample will be a kind of Frege case: an agent will satisfy the antecedent of L, but fail to satisfy the consequent, because the theory of symbols treats tokens as being type identical that are actually causally distinct in the way they function in the different system's cognitive economies. For instance, let R*, being a computational relation between symbols, be a relation such as:

> When S has #it is raining# and #an umbrella is available#, ceteris paribus, S will hold #reach for umbrella#. (Where '#' denotes LOT symbols.)

And suppose that Jose satisfies the antecedent, but likes to be rained on. Hence, he does not satisfy the consequent. This is due to the unique computational role of his #raining# thoughts. Now, it may strike one as building too much into the nature of a symbol to say that Jose's #raining# token is individuated by this feature of its conceptual role. But let us see where *not* doing so leads us. For Frege cases at the computational level cannot be solved. Both Fodor and myself have tried to solve intentional-level Frege cases by saying that they are included in the ceteris paribus

[16] A version of this argument also appears in Schneider (2009b); there, it is used for the purpose of illustrating a problem with molecularism.

clauses of intentional laws, because they are tolerable exceptions (Fodor, 1994; Schneider, 2005). In making such an argument (inter alia), one explains that there was a difference in underlying MOPs (i.e. LOT states), and that the individual didn't realize that the states referred to the same individual. Although it is controversial to cash out MOPs symbolically, it is plausible to expect that such phenomena should be explained at a theoretical level that is sensitive to the particular representations, or MOPs, that the subject has when she has the Frege case. For the proponent of LOT, this level is the symbolic or computational level.[17] Without going into detail concerning the intentional-level cases, let me simply observe that even if this strategy works in the intentional case, the crucial thing to note is that unlike Frege cases arising at the intentional level, in the case of computational-level Frege cases, according to the LOT/CTM picture, there is no theoretically appropriate lower psychological level to appeal to, for, as with intentional level cases, the explanation of the Frege cases involves MOPs. When symbols are typed in a manner that is more coarsely grained than holistic computational role, there will be instances in which individuals satisfy the antecedent, but not the consequent, for reasons that are rightly considered to be psychological and MOP involving, but which LOT *cannot* explain. This incompleteness in the theory is certainly not in keeping with the traditional LOT program, which holds that mental processes are exhaustively syntax driven.

In sum, things do not appear to bode well for positions that run short of the extreme view that symbols are individuated in terms of their total computational roles. Further, this argument is somewhat different from the previous two, as it can be stated in a way that does not assume Classicism or CTM. For the gist of the argument is that without individuation of modes of presentation by total computational (or for non-computational theories, narrow functional) role, then, either will be missed predictions, or there will be counterexamples to certain psychological laws. The argument should generalize to alternate conceptions of guises or modes of presentation. Philosophy of mind has suggested a number of entities that might play the theoretical role of modes of presentation. (Some well-known candidates are: narrow contents, activation patterns in connectionist networks, or LOT symbols.) The basics of the argument seem to present a demand for a taxonomy of psychological states in terms of total functional or computational role, although it is up to the proprietary theory to say how such MOPs are to be construed. The upshot seems to be that, unless one individuates psychological states in this extremely thin manner, either missed predictions or counterexamples will ensue.[18] Thus, my first two arguments, which suggest that symbols would be sliced

---

17 For an in depth discussion of explanation of intentional level Frege cases see Schneider, 2005. Here, I argue that Frege cases are explained by certain intentional generalizations, as well as underlying differences in MOPs. See also Rupert, 2008, for an extension of my approach.

18 This is not to say that *all* psychological kinds need to be individuated in this way. Indeed, this picture is compatible with an appeal to a semantic or intentional level as well, which subsumes thoughts taxonomized by their broad contents. See, e.g. Fodor, 1994; Schneider, 2005.

very thinly, cannot be construed as yielding a psychological kind that is, from the vantage point of any cognitive science, unnecessary and avoidable.

This concludes my discussion of the three arguments. We shall now turn to an important objection to this picture. As noted, there a significant worry that arises for functional role views that define a mental state by the role it plays in one's complete cognitive economy. Critics charge that such views are not 'public'. In the following section, I respond to this worry as it arises for my own view.

2. Publicity-Based Objections

As many know, publicity requirements are common in the concepts literature. There, it is considered by many to be a reasonable requirement on a theory of concept individuation that any plausible theory must provide a sense in which different individuals can have the same concept.[19] Now, in the context of my theory of symbols, both Prinz and Fodor have raised publicity concerns of a related sort.[20] According to the LOT picture, computational generalizations are sensitive to LOT symbols. But I have just claimed that primitive symbols are typed by their total computational roles. And this sort of view is notoriously fine grained, resulting in a situation in which few, if any, symbols are shared. As a result, different individuals will apparently not fall under the same computational generalizations, for, on the symbol processing view of cognition, computational generalizations are supposed to be sensitive to symbol types. At best, psychology would have system specific laws, laws that cover the workings of a particular system at a given time. So psychology will not be public.

The objection is flawed for the following three reasons:

Reason One: Publicity in Terms of Referential/Externalist Explanation

According to the LOT picture, vocabulary items in the language of thought play the role of neo-Fregean modes of presentation (MOPs), being the ways that the individual represents the world. It is crucial to note that LOT doesn't require the view that the MOPs be semantic entities. Rather, LOT symbols, being computational, are plausibly regarded as being non-semantic. Indeed, the LOT view of the nature of MOPs is usually conjoined with an externalist view of semantic content, for many find externalism to be independently plausible. With this in mind, to the extent that there are semantic or intentional generalizations that subsume individuals with respect to broad contents, systems will be able to satisfy

[19] For example, Prinz, 2002 and Fodor, 1998.

[20] In conversation and personal correspondence. Murat Aydede also offers this criticism of a related view of symbol individuation, which he calls the 'Narrow Functionalist Account' in his helpful discussion of LOT symbol individuation (2000). He attributes the narrow functionalist view to Stich, 1983.

the same psychological generalizations despite the fact that technically, they do not share the same MOP/symbol types.

Of course a natural question to ask is: does cognitive science require that there be such generalizations? I believe so. Indeed, Ned Block, Jerry Fodor and Zenon Plyshyn have all argued that cognitive science needs laws that subsume agents by the broadly referential properties of their internal mental states (Block, 1994; Fodor, 1994; Pylyshyn, 1986). Let us call such generalizations 'referential generalizations'. In calling for such generalizations, they argue that such generalizations can capture predictive uniformities between mental states that are distinct ways of representing the same referent. This is because, given that people live in the same environment and have similar mental structures, people's behavior towards the referent tends to converge despite small (and even large) differences in their ways of representing the world. However, if intentional laws are solely sensitive to narrow contents or computational states, any predictive uniformity in their referent-directed behaviors is lost. One may think of water as 'the stuff with that contains both hydrogen and oxygen'; another may think of it as, 'the liquid people like to drink', and so on. Nonetheless, both parties satisfy many of the same water-related generalizations. Different systems having distinct ways of representing the same entity will frequently behave in similar ways because they are embedded in similar environments and because they make similar demands on these environments. But how would this tendency toward similar thoughts and behaviors be captured by the generalizations of a purely narrow psychology?[21]

A critic may retort that this point speaks to intentional explanation, rather than explanation about the internal computational workings of a system. And the original problem is still present for computational explanation, even if shared intentional explanation is possible. For how is *computational* explanation possible if individuals do not share the same symbols? To speak to the computational side of things, I believe that explanation in computational psychology does not literally require shared LOT expressions. Indeed, it has been widely observed that that the main emphasis of computational explanation is *not* the subsumption of events in laws; instead, it is mainly concerned with explanation by functional analysis or decomposition (Block, 1995; Hardcastle, 1996; Cummins, 1983). But as I illustrate in my second reason below, even on a holistic construal of symbols, such explanation can still be public.

[21] The critic may retort that the very need for generalizations sensitive to broad content arises because my theory says that computational states are not public. This is not the case, however. For we should distinguish a situation in which contents/computational states can be individuated in a way that is system-specific, and thus not shared, from a situation in which they are shared, but (what the critic regards as) 'reasonable' cognitive or perceptual differences individuate the mental states. Even if there is a manner of typing of computational states/contents across distinct systems, any plausible theory of typing will distinguish states with very different conceptual roles. However, even when there are 'reasonable differences' between the states, in general, people's behavior towards the referent tends to converge. Such convergence can be captured by broad laws.

Reason Two: Publicity In Terms of Functional Analysis

According to the method of functional analysis, a system is described in terms of the causal organization of its components and the way in which the components interrelate. Those who offer publicity worries fail to note that functional analysis does not require that systems have symbols of the same type. Consider, for instance, one of the most familiar introductory-level discussions of functional decomposition, Ned Block's 'The Mind as the Software of the Brain' (1995). Block provides two examples of functional decomposition; on both of these examples, the description of the machine's functional organization abstracts away from the actual symbols that the machine computes. Consider the first example:

> Suppose one wants to explain how we understand language. Part of the system will recognize individual words. This word-recognizer might be composed of three components, one of which has the task of fetching each incoming word, one at a time, and passing it to a second component. The second component includes a dictionary, i.e. a list of all the words in the vocabulary, together with syntactic and semantic information about each word. This second component compares the target word with words in the vocabulary (perhaps executing many such comparisons simultaneously) until it gets a match. When it finds a match, it sends a signal to a third component whose job it is to retrieve the syntactic and semantic information stored in the dictionary. This speculation about how a model of language understanding works is supposed to illustrate how a cognitive competence can be explained by appeal to simpler cognitive competences, in this case, the simple mechanical operations of fetching and matching (Block, 1995).

Of course, this case is very simple. But it is designed to be like actual cases of functional decomposition in cognitive science. Notice that all of the mental operations are explained without appeal to the particular symbols in the device's memory. This is because different systems will process different words, and may have an entirely different vocabulary in its database. The explanation of fetching and matching tasks needs to abstract away from the particular words, in order to unite similar phenomena into a shared explanatory framework.

Now let us turn to an actual case of functional decomposition in cognitive science. Consider Alan Baddeley's influential explanation of working memory (Baddeley, 1986, 2003). It does not require shared MOPs or symbols because it abstracts away from such details, focusing instead on the general processing of any contents of working memory. Baddeley conceives of working memory (WM) as a system that provides 'temporary storage and manipulation of the information necessary for complex cognitive tasks as language comprehension, learning and reasoning' (1986). His focus is not on the particular contents of memory, but on providing a functional decomposition that begins with a tripartite division of the WM system. Working memory is comprised of a 'central executive', which is an attentional controlling system, and two slave systems: the 'phonological loop', which stores and rehearses speech based information, and the 'visuospatial scratchpad', which holds and manipulates visuospatial information (Baddeley, 1986, 2003). The workings of each of these slave systems are described in a way that abstracts away from the particular items in one's database. Of course, I cannot walk through all the examples of functional decomposition in cognitive science. But it is fair to say that at this point, the burden is on the critic to show that functional decomposition would be ruined by a failure to arrive at a theory of shared symbols. As far as I can tell, even if symbols aren't shared, computational psychology is still public in its most crucial explanatory dimension—that of functional decomposition.

I shall now outline yet another manner in which psychology can be public, even if symbols are not shared. This discussion begins with an objection to the above position.

Reason Three: 'Symbol Neutral' Generalizations

The objection observes the following: although much of scientific psychology is concerned with explaining cognitive capacities, there is also much interest in discovering and confirming laws or effects (Cummins, 2000). And it is in the domain of laws, if not functional decomposition, that the problem with symbol types infects the possibility of computational explanation, or at least an important part of computational explanation. For if symbols are taxonomized by their total computational roles, there cannot be computational laws that cover distinct systems.

Here, I'd like to emphasize that that any failure to arrive at shared symbol types does not infect the possibility of computational explanation in terms of laws. For one thing, computational psychology frequently appeals to generalizations that quantify over LOT symbols, but without quantification over *particular symbol types*. Such statements generalize over LOT expressions without actually specifying them. I shall call these 'symbol neutral' generalizations.

Let me say more. In the context of debates over the plausibility of narrow content, it was suggested that computational psychology actually consists in laws that quantify over MOPs in general, without actually quantifying over particular types.[22] This suggestion strikes me as apt, although, as I will explain below, symbol neutral laws are not the only sorts of laws that computational psychology appeals to. But to focus on the symbol neutral laws, for now, I think there is a strong case to be made that such generalizations are present throughout cognitive science. In the context of the narrow content debate, Block gives the following example:

> If one wants W and also believes that G is required for W, then, ceteris paribus, one will try to do G.

[22] Fodor, 1987. Block also briefly gives this suggestion in his 1998. Block's suggestion concerns narrow contents, but on his view, narrow contents are individuated by computational relations between symbols. I have argued elsewhere that this form of content is not really a form of content at all (Schneider, 2005).

Notice that such generalizations do not require *a theory of* shared MOPs (i.e. narrow contents, LOT symbols, or whatever); all that is required is that there be a principle of individuation that distinguishes MOP types within a given system, that is, synchronically and *intra*personally.[23] Indeed, within fields that have generalizations that cover modes of presentation an appeal to such neutral laws is commonplace. Consider, for instance, generalizations about memory and attention, which abstract away from the particular items being processed, focusing instead on general patterns that explain phenomena like storage capacity, encoding, and retrieval. For example, consider George Miller's generalization about the upper limit on the number of items in working memory—his 'magical number seven, plus or minus two'.[24] And MOP neutral generalizations are rife throughout the literature on concepts as well. Consider, for example, the prototype effects discovered by Eleanor Rosch and her cohorts.[25] MOP neutral generalizations are also found throughout work on social cognition, e.g. in generalizations that are concerned with patterns involving social stereotyping, which abstract away from the details of a particular case, and focus on general patterns of thinking and behavior.[26] And consider 'syntactic' or proof theoretic versions of logical generalizaions like *Modus Ponens* and Conjunction Elimination.

Now, I do not mean to suggest that cognitive science *only* has such generalizations. For computational psychology clearly finds certain more specific generalizations to be of interest. For instance, consider:

(M) The moon looks larger on the horizon than it does in other parts of the sky.

However, I do not think the presence of such generalizations poses a problem for my account of symbols, for such generalizations are best taken as being referential. Intuitively, even if there are such things as shared narrow contents or shared symbols, if this generalization was sensitive to such types, it would artificially exclude individuals who seem to satisfy this generalization. For individuals normally do have differences in the modes of presentation of the moon (or horizon, or sky), while still experiencing the moon illusion.[27]

Indeed, I suspect that it is only in explaining the detailed workings of a particular system that computational psychology needs to appeal to laws that quantify over

[23] Here, I am trying to put the issue in a way that does not assume that researchers in such fields uniformly believe in LOT. Of course, the proponent of LOT will construe MOPs as being symbols. Others will construe MOPs as narrow contents, activation patterns in connectionist networks, etc.

[24] Miller, 1956.

[25] Rosch, 1976 and 1978.

[26] Many such generalizations are laid out in Kunda, 1999.

[27] Some might find my referential interpretation of (M) to be unsatisfactory, as the domain of computational psychology seems to be narrow and syntactic. However, the law, while obviously being a prediction, is actually a statement that is to be explained by a particular computational account of the moon illusion. It is not itself intended to serve as the underlying computational account. For more discussion see Cummins, 2000.

particular symbol types. Such laws would clearly need to quantify over particular symbol types, for they are supposed to detail a machine's transition from one symbolic state to another. If one wanted an account of the particular workings of a system—e.g. to explain how a particular system satisfies a given cognitive function—the explanation would need to invoke particular LOT states. But here I would ask the critics, why would a publicity requirement be appropriate in such contexts? Such explanations, by their very nature, are supposed to deal with the idiosyncratic workings of a particular system.

So my answer to Fodor and Prinz is that a holistic type individuation of primitive symbols does not ruin the possibility of computational explanation for a number of reasons. First, functional analysis is still available. Second, it is fair to say that computational psychology consists in some generalizations that are referential and others which quantify over LOT symbols, but not by their types. Even on a holistic construal of symbols, different individuals are able to satisfy both of these sorts of generalizations. And thirdly, on my view, computational psychology only needs to quantify over particular symbol types when explanation of the detailed workings of a system is being provided. And in this domain, it is far from clear why any publicity requirement is appropriate.

Having set aside publicity worries, we are now ready to turn to an issue requiring more discussion. The three arguments which I presented suggest that symbols should be typed by their roles in computation. But what does total computational role amount to? I would now like to make the notion of a total computational role more explicit, as it is doing a good deal of work in the three arguments. I will also identify a specific individuation condition based on the three arguments.

3. Total Computational Role

In essence, I have claimed that the total computational role of a symbol is the role it plays in the relevant 'program'. But what does this mean? Let us assume that there are internal representations—patterns of energy or matter inside the head—that, according to the symbol manipulation view, fall into symbol types. According to the proponent of LOT, such patterns instantiate rules; items like production rules ('if preconditions 1, 2 and 3 are met, do actions 1 and 2'). We can think of these rules as algorithms, or lines in a larger program; such are the rules that describe cognitive processes. In light of this, the question of what computational role amounts to is really the following question: what rules, (or algorithms) does the brain 'run'?

We can restrict our question a bit: LOT is intended to be a theory that describes cognitive processing; processing in what Fodor calls the 'central' system. By a 'central system' Fodor means a subsystem in the brain in which information from the different sense modalities is integrated, behavior is planned, and conscious deliberation occurs. A central system is informationally unencapsulated: that is, its operations can draw upon information from potentially any cognitive domain (Fodor, 1983). This being said, given that the central systems are supposed to be the domain to which LOT applies, it is natural to look to ongoing and future research on the central systems for the needed algorithms. The point is not which algorithms are sensible candidates—I happen to think certain work is promising, but it is well-known that cognitive science knows far, far more about modular, and in particular, sensory processes than it does about higher cognitive function.[28] An inventory of central algorithms is an undertaking of a future cognitive science.

Nonetheless, we can figure out how to individuate symbols today:

> (CAUSAL) A symbol is defined by the role it plays in the algorithms which describe the central system.

Even if one disagrees with the content of the current host of computational theories of the central systems, it is sufficient to agree with my claim that the central systems will eventually be computationally described—that is to say, there will eventually be algorithms available that can be summoned in an individuation condition in which symbols are typed by their role in the central systems.[29] (And, reader, if you accept LOT yet are not in agreement with the claim that algorithms will be found, it is reasonable to ask: why should you believe in LOT at all? For if the central systems aren't computational, why is LOT, which is a computational account of the mind, supposed to be correct? Isn't the primary domain of LOT supposed to be the central systems?)[30]

We can say a bit more about the content of CAUSAL at this point in time, however. CAUSAL includes at least the following two sorts of algorithms. First, those which I have described as 'symbol neutral generalizations'; that is, generalizations that quantify over LOT symbols in general, but not by their particular types (e.g. consider *Modus Ponens* or George Miller's generalization about the chunks of manageable information that working memory systems can contain). And second, the aforementioned generalizations that quantify over LOT expressions by their specific types (e.g. when system S has the mental sentence #x is in pain# then S has the mental sentence #assist x#). This latter sort of generalization plays a

[28] For a discussion of certain work on the central systems which may be useful in identifying candidate algorithms see Schneider, 2007 and ms.

[29] For an overview of progress on uncovering cognitive processes (e.g. attention and memory) see Gazzaniga, Ivry and Mangun, 2002.

[30] It is worth underscoring that CAUSAL individuates symbols by the algorithms that characterize the central systems and not the actual causal profile of the given symbol. In general, in defining a mental state by its causal role, one can proceed by defining or individuating it only by its actual causal history (what has been called the 'actual causal profile') or one can, more inclusively, define it by its dispositions, some of which may not be exercised. Consider, by way of analogy, choosing between defining a property, like *being glass*, by what it actually happens to do during the time in which it exists, versus defining it by what it can do as well (that is, what it is capable of doing compatible with the laws, under various circumstances). The nature of a piece of glass clearly outruns its actual causal profile. Similarly, if symbols are individuated only by the algorithms they happen to satisfy (analogously, the lines of a program that actually run) they will be mischaracterized as being type identical in cases in which the actual history of the device is too short.

key role in symbol individuation. For if only the first sort of generalization typed symbols, as Murat Aydede has noted, it would not be sufficient to distinguish intuitively distinct LOT expressions, as many distinct symbol types which have the same grammatical role (e.g. #dog#, #cat#) can satisfy the same symbol specific generalizations; such generalizations are quite generic.[31]

We have yet an important issue to consider concerning how to formulate CAUSAL. Does CAUSAL merely classify symbol types *within a given system*, or, in addition to this, does it license claims about the 'across system' case? That is, does it allow that different tokens in distinct systems can be type identical insofar as they play the same total computational role? Throughout this paper I have presupposed an affirmative answer; but it is important to illustrate why this is the case. To answer our question we must ask whether the three aforementioned arguments merely suggest a 'within system' condition, or whether they speak to the across system case as well. To appreciate the difference consider:

P1. Within a given system, two tokens belong to the same type if and only if they play the same total computational role.
P2. For any two systems, or within a given system, two tokens are the same type if and only if they have the same total computational role.

(Where, again, by 'total computational role' I mean the role the symbol plays in the algorithms that cognitive science ultimately employs to characterize the central systems.) Note that P1 is weaker than P2 in the sense that P1 is silent concerning the across system case. P1 is compatible with the denial that any across system condition exists; on the other hand, P1 is also compatible with the following:

P3. For any token x in system S1 and any token y in S2, x and y can be of the same type yet play different computational roles.

In this case, there is a 'within system' condition, P1, that types tokens by the total computational role the symbol plays in the given system. But for the case of typing tokens across different systems, a different standard applies (perhaps, for instance, molecularism holds).

Now, what principle(s) do the three arguments support? Argument Two supports P1 only, for it was clearly limited to the within system case (for it began with, 'Assume that *within a given system,* two primitive symbol tokens, a and b, are of symbol type T1, but a has the following causal relation that b does not have...').

[31] Aydede, 2000. As Aydede notes, it is the inclusion of these sorts of generalizations as algorithms which individuate symbols which leads to the failure of symbols to be shared. If only the former sorts of generalizations individuated symbol types, symbols would be shared all the time. Because this latter sort of generalization contains types that aren't shared some may suggest that such are not bone fide generalizations. However, they are generalizations in the following senses: (i), as with the first sort of generalization, they are rules the system uses to state abstract relationships between variables, allowing one to express generalizations compactly, learn, and represent relationships that hold for all members of a given class. (Marcus, 2001 p. 5). (ii). They can be shared, at least in principle, by systems having all and only the same central algorithms.

What about Argument One? Overall, the discussion in Argument One refers to individuation within the rules of a particular program. However, the argument could be construed as supporting P2, the view that tokens, even in distinct systems, must function the same way in the same program to be type identical. Consider, for instance, the example of a game of chess, which Haugeland raised as an example of a symbol manipulation game. It is very much in keeping with this view to say that different tokens on different boards are type identical (e.g. tokens of the type, *rooks*) insofar as they function the same way in all and only the same rules.

We do not need to develop this issue however, for we find further support for P2 from Argument Three. As discussed, there is at least a position in logical space in which P1, being silent about the cross-system case, is conjoined with a cross system condition like P3. However, Argument Three can be viewed as both an argument against P3 as well as an argument for P1 and P2. This argument says that psychology needs a natural kind that is individuated by the total computational role, otherwise there will be counterexamples or missed predictions. Recall that the argument involved stipulating that two tokens, x and y, are individuated by the same condition, CR, yet x has a computational relation, R*, that y lacks. As it happens, the argument can actually be presented in two ways: x and y could be within a given system or in distinct systems. So consider the distinct system case, and as before, assume that x has a computational relation, R*, that y does not, and that further, the individuation condition, CR, is insensitive to R*. On the common assumption that causal relations are backed by laws, there would then be a computational law, or at least a nomologically necessary generalization, L, which specifies R*. But then either: (i), both *a* and *b* will *not* be subsumable in L. Or, (ii), they will both be subsumable in L. In the case of (i), the theory will have a missed prediction: it will miss that *a* has a causal relation that is specified by L. (ii) Now consider the second scenario, in which they will both be subsumable in L. In this case *b* does not have the causal relation detailed by L. So we wouldn't expect it to behave in accordance with L. Hence, *b* will be a counterexample to L. This modified version of Argument Three thus provides reason to believe that across systems, tokens are type identical only if they are characterized by all and only the same algorithms. And because it is uncontroversial that sameness of total computational role is sufficient for sameness of type, the condition can read 'if and only if'. So Argument Three yields, in its across system incarnation, an argument against P3 and for P2. (And in its within system incarnation, it is an argument for P1.)

A further problem with P3 arises as well. For let us ask: what would an across-system symbol type really amount to? Unless one can refute the three arguments one is committed to type individuating vocabulary items in the language of thought within a given system by total computational role. So if P3 is appealed to as well, this leaves one with two sorts of classificatory schemes for vocabulary items in the language of thought: one scheme for the within system case and another for the cross system case. Such an approach is on the wrong track, however: it does not make sense to speak of a 'language of thought' that is only supposed to apply to the across system case, and yet be the language of *thought*. For a language to qualify as a language of thought, it should be the actual language that the system computes in, and which describes the system's underlying psychological processes.[32]

The upshot: the arguments of this paper suggest that P1 and P2 hold but P3 is false—an across system condition is licensed only insofar as the systems in question are characterized by all and only the same central algorithms. And this brings us to the important issue of why, given that P2 holds, I've needed to argue that publicity obtains even if symbols aren't shared. The reason is that, as noted, given that P2 is correct, for shared symbols to exist at all, different individuals' central systems must be characterized by all and only the same algorithms. Plausibly, this occurs in the Twin Earth case, as the 'twins' are molecular duplicates. (And indeed, this could be an explanatory advantage for LOT, if it adopts the present view of symbol individuation, for it allows for sameness in symbol/MOP across twins, despite semantic differences.) The problem concerns the ordinary (non-twin) case. Research on the central systems is only in its infancy; to say that different individuals' cognitive systems will be specified by the same algorithms strikes me as highly speculative, especially in light of the fact that there are massive individual variations in prefrontal cortical processing. The plausibility of LOT's theory of symbols should not stand or fall with this issue. It is for this reason that I assumed symbol types are not shared, arguing that explanation is nonetheless 'public'.

Let me now turn to some important objections to this notion of total computational role and to the related individuation condition. First, a natural objection is that my appeal to as yet undiscovered algorithms is problematic, because it is merely a promissory note to a future cognitive science. First, the fact that the nature of the algorithms is largely unknown does not make the three arguments offered in section one any less plausible. For these arguments would seem to apply, whatever the relevant algorithms turn out to be. Second, such an objection would also rule out all forms of *a posteriori* functionalism, as all appeal to largely as yet undiscovered laws, but without any substantial argument as to why such *a posteriori* approaches are implausible. For the present approach is merely an instance of the general approach of psychofunctionalism. This would leave us with purely armchair approaches to mental state individuation, yet without any real argument for the ineffectiveness of a posteriori functionalism. And third, rejecting CAUSAL for this reason is tantamount to ruling out all proposals that individuate an entity by causal powers, (e.g. Sydney Shoemaker's theory of the nature of properties), for presumably, such causal powers are ultimately a matter of scientific discovery. Clearly, one can offer compelling arguments for an individuation condition, say,

[32] It could be that there is some incentive for the proponent of LOT to adopt an intrapersonal individuation condition for symbols *and* an across system MOP of a different sort (where such are not LOT symbols). While the paper doesn't rule out such a project, I see no reason to believe that psychology needs such a kind. For even if individuation by total computational role is taken to be the only manner of individuation of MOPs, (as I am happy to assume herein), I have argued that psychology can, in fact, be public.

on the nature of properties, in absence of knowledge of the precise content of the laws, (dispositions, causal powers, etc.) themselves.

A second objection will occur to those familiar with Jerry Fodor's attack on the central systems. Fodor himself would deny that the program can be specified. Stronger yet, he would be inclined to deny that such a program exists. For interestingly, since the *Modularity of Mind* he has held that the central systems seem to defy computational explanation. (Indeed, this is a central claim of both his *The Mind Doesn't Work that Way* (2000) and *LOT2* (2008).) In the context of our discussion, Fodor's well-known concerns, if correct, as well as challenging the computational nature of the central systems, present an additional obstacle to the LOT program. For if symbols must be individuated by their computational roles, and if Fodor is correct that the central systems are probably not even computational, the language of thought program *must* fail. For how, in such a scenario, can symbols be individuated by their *computational* roles?[33] If Fodor is correct, the central systems are not computational to begin with. This objection is particularly worrisome (and indeed, surprising) as it comes from the leading proponent of LOT and CTM.

The problems that Fodor worries plague CTM divide into two kinds, and both purport to show that the success of cognitive science will likely be limited to the modules. The first sort of problem concerns what Fodor has called 'global properties'; features that a mental sentence has which depend on how the sentence interacts with a larger plan (i.e. set of sentences), rather than the type identity of the sentence alone. In a key passage, Fodor explains:

> The thought that there will be no wind tomorrow significantly complicates your arrangements if you had intended to sail to Chicago, but not if your plan was to fly, drive or walk there. But, of course the syntax of the mental representation that expresses the thought #no wind tomorrow# is the same whichever plan you add it to. The long and short is: the complexity of a thought is not intrinsic; it depends on the context. But the syntax of a representation is one of its essential properties and so doesn't change when the representation is transported from one context to another. So how could the simplicity of a thought supervene on its syntax? As please recall, CTM requires it to do (2000, p. 26).

The rough argument (which I shall call 'The Globality Problem') seems to be the following. Cognition seems sensitive to global properties. E.g. the addition of a new sentence in LOT frequently complicates an existing plan. But CTM holds that cognition, being computation, is sensitive only to the 'syntax' of mental representations. And further, syntactic properties are *context insensitive* properties of a mental representation. That is, what a mental representation's syntactic properties are does not depend on what other mental representations in a plan it is combined with: it depends on the type identity of the LOT sentence. But whether a given

[33] Notice that this observation holds for the aforementioned molecularist proposal as well.

mental representation has the global properties that it has will typically depend upon the *context* of other representations in a plan. That is, it depends upon the nature of the other LOT sentences in the relevant group. So it seems that cognition then cannot be wholly explained in terms of computations defined over syntactic properties. Thus, CTM is false.[34]

The second problem concerns what many have called, 'The Relevance Problem': the problem of whether and how humans determine what is relevant in a computational manner. The Relevance Problem is often put in the following way: If one wants to get a machine to determine what is relevant, it seems that the machine would need to walk through virtually every item in its database, in order to determine whether a given item is relevant or not. This is an enormous computational task, and it could not be accomplished in a quick enough way for a system to act in real time. Of course, humans make quick decisions about relevance all the time. So, it looks like human domain general thought (i.e. the processing of the central systems) is not computational (Fodor 2000, 2008).

Elsewhere, I have responded to both of Fodor's concerns. First, Kirk Ludwig and myself have argued that the problem that Fodor believes global properties pose for CTM is a non-problem (Ludwig and Schneider, 2008; Schneider, 2007). Further, I've argued that although the relevance problem is a serious research issue, it does not justify the grim view that cognitive science, and CTM in particular, will likely fail to explain cognition (Schneider, 2007). While I do not have time to delve into all of the considerations raised against Fodor's problems, I can quickly sketch a few short arguments.

Suppose that one can show that both problems can emerge in the context of uncontroversially computational processes. Then, the presence of a globality or relevance problem does not entail that the system in question is non-computational. I shall now proceed to do this. Consider a tinker toy chess-playing program. Suppose that a human opponent makes the first move of the game, moving a certain pawn one square forward. Now, the program needs to decide, given the information of what the previous move was, which future move to execute.

(i) The Globality Problem Emerges. Suppose that there are two game strategies/plans in the program's database, and the program needs to select one, given the first move. Let one plan involve getting the bishop out early in the game, while the other plan involves getting the rook out early in the game. (Where 'early' means, say, within three turns.) Now, let us formulate a sort of globality worry: notice that the impact that the addition of the information about what the opponent's first move was on the simplicity of each of the two plans does not appear to supervene on the type identity of the string of symbols encoding the information about the opponent's first move. Instead, the impact of the addition of the string of symbols to the simplicity of each plan depends on the way that the string interacts with the other sentences (i.e. syntactic strings) in the plan. Thus, (the

[34] For a longer discussion of this issue and a more detailed version of Fodor's argument see Ludwig and Schneider, 2008.

Globality Argument continues) the processing of the chess program is not syntactic, and hence, not computational. So, it seems that a Globality Problem emerges in the context of highly domain specific computing (Schneider, 2007, 2009a).

(ii) The Relevance Problem Emerges. Skillful chess playing involves the ability to select a move based on the projected outcome of the move as far into the future of the game as possible. So chess programmers routinely deal with a massive combinatorial explosion. In order to quickly determine the best move, clever heuristics must be used. This is precisely the issue of locating algorithms that best allow for the quick selection of a future move from the greatest possible projection of potential future configurations of the board (Marsland and Schaeffer, 1990). And this is just the Relevance Problem, as Fodor and other philosophers have articulated it (Schneider, 2007, 2009a).

In sum: both problems emerge at the level of relatively simple, modular, and uncontroversially computational processes. If both problems can occur in the context of uncontroversially computational processes, the presence of a globality or relevance problem does not entail the conclusion that the system in question is non-computational. And this is the conclusion which is needed to undermine CAUSAL.

Further, we can quickly identify the underlying flaw in the Globality Argument. The globality problem is supposed to arise from the fact that the same LOT sentence, e.g. #no wind tomorrow#, may differ in the effect it has, depending upon the type identity of the other sentences in the plan. However, this fact does not really introduce a problem for CTM, for it is compatible with the requirement that LOT syntax be context insensitive (i.e. the requirement that tokens of the same symbol type will make the same syntactic contribution to every belief set that they figure in). The same mental sentence can do this because all a LOT sentence contributes to a computation is its type identity; the type identity of a sentence can have a different impact on different plans/groups of sentences. The impact depends upon the type identity of the added sentence, together with the nature of the algorithms and the type identity of the other sentences in the group. To consider an analogous case, consider the case in which one adds a new premise to an existing argument in first-order logic. When the premise is put into a different argument, the same premise may have a different impact; for instance, in one case, it may bring about a contradiction, in another case, it may not. But the difference in impact, although it is not a matter of the type identity of the premise alone, is still syntactic, depending on the type identity of the premise, together with the type identity of the other sentences in the argument, and the rules (Schneider, 2007).

In contrast to the globality problem, which is merely a non-problem, the relevance program *does* present quite a challenge to programmers; the challenge is to select judicious algorithms which maximize the amount of information subject to the constraints of real time. However, if my above argument concerning relevance is correct, it is implausible to claim that a relevance problem entails that the system in question is likely non-computational.[35] Now, there might be an alternative, more plausible formulation of the problem that relevance presents for CTM; in Schneider (2007) I walk through different formulations that could lend support to Fodor's view that the central systems are likely non-computational. But for now, let me suggest that a very different way to proceed with respect to the Relevance Problem is to assume that the presence of a human relevance problem is not terribly different from relevance problems existing for other *computational* systems. But, in the human case, the 'solution' is a matter of empirical investigation of the underlying brain mechanisms involving human searches. This alternative approach assumes that evolution has provided *homo sapiens* with algorithms that enable quick determination of what is relevant, and further, it is the job of cognitive science to discover the algorithms (Schneider, 2007). On this view, we must resist Fodor's suggestion that research in cognitive science should rest at the modules (Fodor, 2000). It seems then that individuation by computational role, and thus LOT, are still in business.

4. Conclusion: Reconfiguring the LOT Approach

We've covered a good deal of terrain. I have supplied three arguments for the individuation of symbols by total computational role. I then argued that psychological explanation can be public, even on a holistic construal of symbols. I then specified the notion of computational role of interest. My overarching view is that the present theory of symbols, while untraditional in the sense that different individuals do not share symbols, deserves further consideration. Further, if any of the three arguments for symbol holism hold, then it is fair to say that this view of symbols is the *only* theory that the proponent of CTM can appeal to.

Now, indulge me for a moment and assume that my three arguments work. Under this assumption, LOT looks very different. Remember: LOT was developed in the absence of a theory of symbols, despite the ironic fact that its key contention is that cognition is *symbol* processing. Now, given the centrality of symbols to

[35] Might Fodor's concern be instead that the central systems cannot be computational because computational systems cannot solve relevance problems? But this is incorrect; there are already programs that carry out domain general searches over vast databases. Consider Internet search engines. In about 200 ms. one will receive an answer to a search query involving two apparently unrelated words that involved searching a database of over a billion webpages. Is his idea, instead, that there is something distinctive about searches that involve the central systems, making such searches infeasible for the central systems, and thus suggesting that the central systems are not computational? If so, I'm not sure how the argument is supposed to go: Fodor's Relevance Problem concerned how to sift through masses of data in real time. But domain generality entails nothing about the size of a database that a search draws from. Consider, e.g. a database recording the mass of every mass-bearing particle in the universe. This would be domain specific, yet be of a much greater size than any search that a central system undergoes, for it involves far more information than is encoded in any human's memory (Schneider, 2007).

LOT, it is not surprising that when the dust finally settles on the question of what symbols are, the very face of the LOT program is altered. While certain features of the reconfigured LOT will be controversial, there are some clear improvements to the LOT framework. These are the following. If my view of symbols is correct, then LOT has a concrete theory of the fundamental nature of cognitive mental states (or 'MOPs'), and further, it can summon these MOPs in an account to naturalize intentionality, explaining how intentional mental states are ultimately physical relations between the symbolic mind and entities in the world. In addition, once the algorithms describing higher-cognition are well understood, a taxonomy of symbolic types can be provided, a key step to determining whether symbolic computations are in fact realized by connectionist networks or whatever structures are appealed to by a penultimate computational neuroscience.

Other results of the present theory of symbols will likely be more controversial. First, if it is correct, as discussed, the central systems *must* be computational. This is because symbols themselves are individuated by central algorithms. *So the success of LOT requires that Fodor's injunction that research in cognitive science rest at the modules be resisted* (Fodor, 2004; Ludwig and Schneider, 2008; Schneider, 2007). But again, I am unworried by this, for I've argued that Fodor's arguments that the central systems are likely non-computational are flawed.

A second result is even more controversial. If my view of symbols is correct, the LOT program will feature an entirely different, and arguably superior, version of Conceptual Atomism (also called 'Informational Atomism'). Conceptual Atomism holds that lexical concepts lack semantic structure, being in this sense 'atoms'. It further holds that a concept is individuated by two components: its broad content and its symbol type. Because, as discussed, symbol natures have been neglected, only the semantic dimension of the theory seems to have been developed. (Indeed, the Conceptual Atomists' concepts are often taken as being *equivalent* to broad contents, despite the fact that they are individuated by their symbol types as well).[36] Now, in the literature on concept individuation the LOT program 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, p. 34). Indeed, Fodor proclaims in *Concepts* that pragmatism is a '. . . catastrophe of analytic philosophy of language and philosophy of mind in the last half of the twentieth century'. And in his *LOT 2* he sees the development of conceptual atomism, and indeed, of LOT itself, as representing an important alternative to pragmatism: '. . . [O]ne of the ways LOT 2 differs from LOT 1 is in the single-mindedness with which it identifies pragmatism as the enemy *par excellence* of Cartesian realism about mental states' (2008, p. 12).

[36] For discussion of Conceptual Atomism see (Fodor, 1998; Laurence and Margolis, 1999; Levine and Bickhard, 1999).

Just as LOT is supposed to be non-pragmatist, so too LOT is said to be *Cartesian*. For instance consider a typical characterization of the Pragmatist/Cartesian opposition appearing in *Philosophical Studies*, by Bradley Rives:

> Pragmatists claim that concepts are individuated in terms of the role they play in the cognitive lives of thinkers, e.g. in terms of their role in inference, perception, and judgment. Cartesians, on the other hand, hold that *none* of the epistemic properties of concepts are concept-constitutive (Rives, 2009).

Rives calls me a 'Cartesian' (p. 23)—this is doubtless guilt by association.

But LOT isn't Cartesian; it cannot be. If I am correct about the nature of symbols then *both LOT and the related doctrine of Conceptual Atomism are pragmatist theories*. Consider: if Concept Pragmatism is, as Fodor claims, the view that a concept's nature is, at least in part, a matter of the role it plays in one's mental life, then Conceptual Atomism must embrace pragmatism. For I've argued that symbols must be individuated by the role they play in one's cognitive economy. Now, perhaps concepts shouldn't be individuated by symbol types. However, the proponents of LOT are likely to say they that concepts are thus individuated, and to this extent they are thereby committed to concept pragmatism. Further, even if the proponent of LOT sets aside Conceptual Atomism and does not individuate concepts by symbol types, it is still inevitable that LOT is pragmatist. For symbols, individuated by the role they play in thought, are the very soul of LOT. The appeal to pragmatism is thereby inescapable.

One final observation. Some might have the nagging worry that my results actually leave the dialectical situation worse for LOT. For the fact that symbols are not shared, from system to system, means that overall, LOT is less attractive than the other competing theories of the inner vehicle of thought. I am not so convinced that there are plausible non-holistic theories of narrow content, (of a sort that are plausibly invoked as kinds in cognitive science), but I must leave this issue for another time.[37] However, it is fairly uncontroversial that connectionist state individuation is holistic.[38] And, in any case, argument three offered a general challenge to other theories of MOP individuation; unless MOP types are cut very finely there will be counterexamples to psychological laws or psychology will be incomplete. If this argument is correct, then the general quest for shared MOPs may be leading us in the wrong theoretical direction.

Additionally, it is crucial to bear in mind the motivations for advancing a criterion for individuating symbols which were raised in the course of our discussion:

(i) That CTM, as a computational theory, provide a well defined notion of a symbol;

---

37 I discuss this issue further at Schneider, 2005. For nice discussions of the various theories of narrow content see Prinz, 2002, and Segal, 2000.

38 Indeed, the chief proponent seems to agree (Churchland, 2005).

(ii) That LOT specify the nature of MOPs, which are supposed to be the inner vehicle of thought and which are supposed to figure in an account of how to naturalize intentionality;
(iii) To express the intuitive sense in which twins in the Twin Earth case share the same inner psychological states;
(iv) To provide a kind which enables psychology to detail the computational configuration of a particular system, and to explain the narrow causation of thought.

In sum, the theory serves a variety of important functions. And crucially, although symbols aren't shared, there is, surprisingly, no violation of 'publicity'. For I have observed that there are numerous ways in which computational psychology is still public, even if symbol holism is in force.

Department of Philosophy,
Center for Cognitive Neuroscience
Institute for Research in Cognitive Science
The University of Pennsylvania

References

Aryo, D. 1996: Sticking up for Oedipus: Fodor on intentional generalizations and broad content. *Mind & Language*, 11, 231–45.

Aydede, M. 2000: On the type/token relation of mental representations. *Facta Philosophica. International Journal for Contemporary Philosophy*, 2, 23–49.

Aydede, M. and Robbins, P. 2001: Are Frege cases exceptions to intentional generalizations? *The Canadian Journal of Philosophy*, 31, 1–22.

Baddeley, A. 1986: *Working Memory*. Oxford: Clarendon Press.

Baddeley, A. 2003: Working memory: looking backward and looking forward. *Nature Reviews Neuroscience*, 4, 829–39.

Block, N. 1994: Advertisement for a semantics for psychology. In S. Stich and E. Warfield (eds), *Mental Representations; A Reader*. Cambridge: Blackwell.

Block, N. 1995: The mind as the software of the brain. In D. Osherson, L. Gleitman, S. Kosslyn, E. Smith and S. Sternberg (eds), *An Invitation to Cognitive Science*. Cambridge, MA: MIT Press.

Block, N. 1998: Holism, mental and semantic. In E. Craig and L. Floridi (eds), *The Routledge Encylopedia of Philosophy*. New York: Routledge.

Churchland, P. 2005: Functionalism at forty, a critical retrospective. *The Journal of Philosophy*, 102, 33–50.

Cummins, R. 2000: How does it work? Versus what are the laws? Two conceptions of psychological explanation. In F. Keil and R. Wilson (eds), *Explanation and Cognition*. Cambridge, MA: MIT Press, 114–44.

Cummins, R. 1983: *The Nature of Psychological Explanation*, Cambridge, MA: MIT Press.

Devitt, M. 1991: Why Fodor can't have it both ways. In B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and His Critics*. Oxford: Blackwell.

Fodor, J. A. 1987: *Psychosemantics*. Cambridge, MA: MIT Press/A Bradford Book.

Fodor, J. A. 1994: *The Elm and the Expert: Mentalese and its Semantics*. Cambridge, MA: MIT Press.

Fodor, J. A. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Fodor, J. A. 2000: *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.

Fodor, J. A. 2004: Having concepts: a brief refutation of the twentieth century, *Mind & Language*, 19, 29–47.

Fodor, J. A. 2008: *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.

Fodor, J. and LePore, E. 1992: *Holism: A Shoppers Guide*. Oxford: Blackwell.

Fodor, J. and Pylyshyn, Z. 1988: Connectionism and cognitive architecture: a critical analysis. *Cognition*, 28, 3–71.

Gazzaniga, M., Ivry, R. and Mangun, G., 2002: *Cognitive Neuroscience*, 2nd edn. New York: W.W. Norton and Co.

Hardcastle 1996: *How to Build a Theory in Cognitive Science*. New York: SUNY Press.

Haugeland, J. 1985: *AI: The Very Idea*. Cambridge, MA: MIT Press.

Laurence, S. and Margolis, E. 1999: Concepts and cognitive science. In E. Margolis and S. Laurence (eds), *Concepts: Core Readings*. Cambridge, MA: MIT Press, 3–81.

Levine, A. and Bickhard, M. 1999: Concepts: where Fodor went wrong. *Philosophical Psychology*, 12, 1, 5–23.

Ludwig, K. and Schneider, S. 2008: Fodor's critique of the classical computational theory of mind. *Mind & Language*, 23, 123–43.

Marcus, G. 2001: *The Algebraic Mind*. Cambridge, MA: MIT Press.

Marsland, A. and Schaeffer, J. (eds) 1990: *Computers, Chess, and Cognition*. New York: Springer-Verlag.

Miller, G. 1956: The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review*, 63, 81–97.

Pessin, A. 1995: Mentalese syntax: between a rock and two hard places, *Philosophical Studies*, 78, 33–53.

Prinz, J, 2002: *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

Pylyshyn, Z. 1986: *Computation and Cognition*. Cambridge, MA: MIT Press.

Kunda, Z. 1999: *Social Cognition: Making Sense of People*. Cambridge, MA: MIT Press.

Rives, B. 2009: Concept Cartesianism, concept pragmatism, and Frege cases. *Philosophical Studies*, 144, 211–38.

Rosch, E., 1976: Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439.

Rosch, E. 1978 *Principles of Categorization*. In E. Rosch and B. B. Lloyd (eds), *Cognition and Categorization*. Hillsdale, NJ: Erlbaum. Reprinted in: Margolis, E. and Laurence, S. (eds), 1999: *Concepts: Core Readings*. Cambridge, MA: MIT Press.

Rupert, R. 2008: Frege's puzzle and Frege cases: defending a quasi-syntactic solution. *Cognitive Systems Research*, 9, 76–91.

Schneider, S. 2007: Yes, it does: a diatribe on Jerry Fodor's *The Mind Doesn't Work That Way*. *Psyche*, 13, 1–15.

Schneider, S. 2005: Direct reference, psychological explanation, and Frege cases, *Mind & Language*, 20, 423–47.

Schneider, S. 2009a: The language of thought. In P. Calvo and J. Symons (eds), *Routledge Companion to Philosophy of Psychology*. New York: Routledge.

Schneider, S. 2009b: LOT, CTM, and the elephant in the room. *Synthese*, 170, 235–50.

Schneider, S. ms.: The central systems as a computational engine.

Segal, G. 2000: *A Slim Book on Narrow Content*. Cambridge MA: MIT Press.

Stich, S. 1983: *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, MA: MIT Press.

Stich, S. 1993: Narrow content meets fat syntax. In B. Loewer and G. Rey (eds), *Meaning in Mind. Fodor and his Critics*. Oxford: Blackwell.

Wermter, S. and Sun, R. (eds) 2000: *Hybrid Neural Systems*. New York: Springer.

Wilson, R. 1997: *Cartesian Psychology and Physical Minds: Individualism and the Science of the Mind*. Cambridge: Cambridge University Press.