The Empirical Investigation of Mental Representations and Their Composition

Joe Thompson

Simon Fraser University

Abstract: The postulation of compositional mental representations, and computational neural systems that are sensitive to them, is so central to cognitive psychology that it seems, for all practical purposes, insulated from revision. According to this view, which I identify with the Language of Thought (or LOT) hypothesis, mental representations must possess syntactic structure because only the purely physical, syntactic, features of mental representations could conceivably fit into causal explanation. It is not contentious that the ultimate success or failure of the LOT hypothesis will lie in its capacity to provide explanations, predictions, and direction to research. What is needed, however, is a non-tendentious method for evaluating how the LOT hypothesis fares on these criteria. The best way to evaluate the LOT hypothesis, it seems, would be to formulate research projects whose consequences bear directly on its plausibility. The difficulty is that such research projects seem hard to design. I will argue that if we are willing to adopt some form of construct validity theory, then such research projects already exist. They might be used to reject the LOT hypothesis, or perhaps revise our views on the syntactic structure of mental representations.

Keywords: Mental Representations; Language of Thought; Syntactic Composition; Basal Ganglia; Recursive Embedding

1: The Language of Thought Hypothesis

1.1: Mental Representations

The Language of Thought hypothesis (or LOT hypothesis) is thought to be implicit in many theories of neural architecture. The hypothesis calls for structured entities in the brain that represent the world, and the syntactic composition of these mental representations determines their expressive and causal powers.[i]

It seems difficult to craft a research project with the *specific* goal of evaluating the LOT hypothesis. This is not to say that claims about compositional mental representations are inherently untestable, or in danger of being unfalsifiable. But the belief in compositional mental representations is so central to cognitive psychology that it seems, for all practical purposes, *insulated* from revision. Throwing out the Language of Thought, at least for those already working under its purview, would be tantamount to a paradigm shift.

The relative insulation of the LOT hypothesis might explain why it is often evaluated without reference to evidence. Fodor (1981), for example, argued that compositional mental representations are the ally of belief-desire psychology, because they provide an intuitive explanation for how the *content* of a belief could cause behaviour. Objections to the coherence of mental representations, along the lines of private language arguments (Wittgenstein, 1959), are similarly non-empirical.

While we need not object to a priori argumentation in science, we should be realistic about the prospects for winning over our theoretical competitors. It is not contentious that the ultimate success or failure of the LOT hypothesis will lie in its capacity to provide explanations, predictions, and direction to research. What is needed, however, is *a non-tendentious* method for evaluating how the LOT hypothesis fares on these criteria. The best way to evaluate the LOT hypothesis, it seems, would be to formulate research projects outcomes consequences bear directly on its plausibility. The difficulty is that such research projects seem hard to design. I will argue that such research projects do exist, and it may be that they can be used to reject the LOT hypothesis, or at least revise our views on the composition of mental representations.[ii]

1.2: Composition and LOT

According to the computational-representational theory of mind,[iii] compositional mental representations have causal powers because the mind's computational architecture is sensitive to the syntactic structure of its mental representations. This notion of COMPOSITION is central to the LOT hypothesis.[iv]

Of course, the syntax and the semantics of 'mentalese' must be suitably related if this account is going to work. Ideally, semantic composition would be definable along the same rules that capture the syntactic composition. This would allow physical systems to track semantic content in the same way that a proof checker can track inference, at least in those cases where organisms are supposed to track content.[v] Consider, for instance, the causal/inferential powers that Fodor (2008) bestows upon the content of belief ``granny left and auntie stayed."

> That the logical syntax of the thought is conjunctive (partially) determines, on the one hand, its truth-conditions and its behavior in inference and, on the other hand, its causal/computational role in mental processes. I think that this bringing of logic and logical syntax together with a theory of mental processes is the foundation of our cognitive science […] (Fodor, 2008, p. 21).

The principle distinction between a theory positing mental representations and the Language of Thought hypothesis, then, lies in the notion of COMPOSITION.[vi] The idea is that there could be mental representations that are not compositional. The content of these representations would be determined by the contents of their atoms alone, and not their structural features. A radical associationist, for example, might explain behaviour by appealing only to associative pairings between structurally atomic ideas (for a discussion of radical associationism, see Rey, 1997). Such a view could allow for mental representations without taking up the LOT hypothesis.

2: LOT and Language Science

LOT theorist's, for reasons to be discussed, will be sympathetic to claims about the recursive structure of mental representations. My worry, however, is that such claims are problematic on empirical grounds.

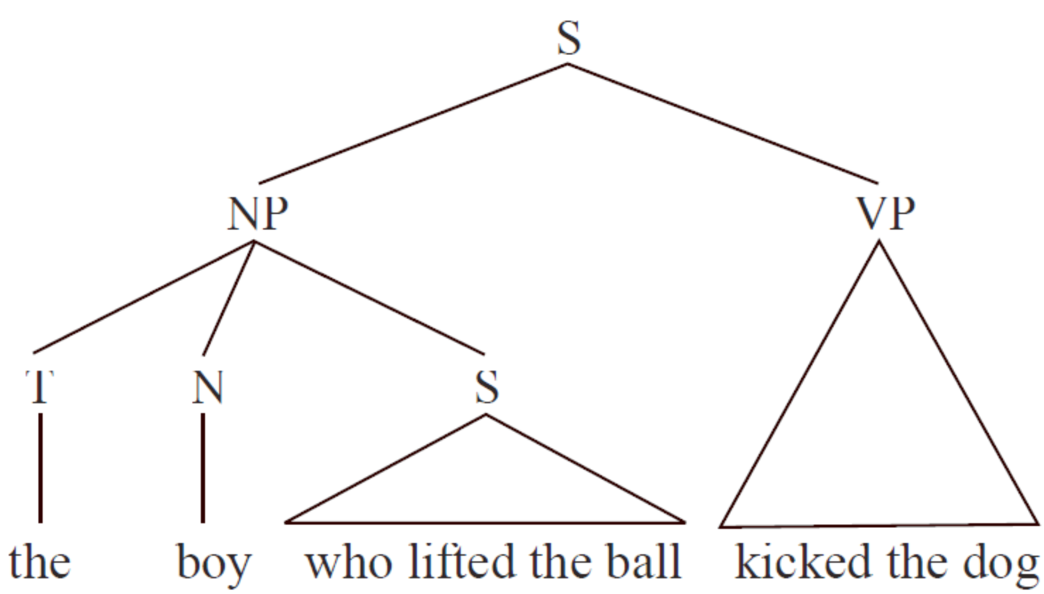
One might try to defend claims about the structural complexity of a natural language by arguing that adequate formal representation of a language requires a language of certain complexity. Chomsky & Miller (1963) argued, for instance, that natural language exhibits *hierarchical complexity*, which is supposed to be apparent in relative clauses where the subject is separated from the verb phrase by a centre embedded clause. For example,

> *The following conversation*, which took place between the two friends in the pump-room one morning, after an acquaintance of eight or nine days, *is given as a specimen of their very warm attachment* […] (emphasis added, Austen, 1917, p. 23)

One can see how the long-distance dependencies apparent in such clauses pose a sequencing problem for organisms (Hauser et al, 2002). At the very least, understanding the sentence seems to require sensitivity to the relationship between the main verb and the subject, despite the distance between them.

Phrase-structure grammars have traditionally been used to provide structural descriptions for these sentences. Figure 1 constitutes a simple example.[vii]

Figure 1: A recursively centre-embedded sentence.



While the LOT theorist will happily accept that English sentences contain noun phrases and verb phrases, it is not obvious that this phrase-structure tree could be read as describing *the structure of a mental representation*. It is unclear, for instance, whether a formal representation of the syntax of mentalese would include noun phrases and verb phrases as grammatical categories.[viii]

Fortunately, this stronger claim about the syntax of representations is not important for my argument. Here I am only interested in whether mental representations contain *recursive elements*. In phrase-structure grammars a grammatical category (or *non-terminal symbol* such as NP, or VP), is a recursive element if and only if the symbol is dominated by another instance of the category further up the tree. A recursive element is also centre-embedded if it occurs in neither the left-most branch, nor the right-most branch, of the tree. The symbol S (which stands for 'sentence') in figure 1 is a recursive centre-embedded sentence. In what follows I will argue that empirical evidence will press the LOT theorist to accept

> (1) Some mentalese sentences contain centre-embedded recursive elements, and perhaps other mentalese sentences, as constituent parts.

(1) should seem plausible to the LOT theorist. This is because long-distance dependencies, so the story goes, present an insurmountable problem for radical associationism, one of the LOT theorist's most established competitors.

Radical associationism is the view that linguistic behaviour can be explained with respect to association-relations between ideas, stimuli, or responses without positing *compositional* mental representations. The problem with such views, according to Fodor (2008), is that they tend to assume that the associations are formed on the basis of temporal-contiguity.

> […] Hume held that ideas became associated as a function of the temporal contiguity of their tokenings. […] Likewise, according to Skinnerian theory, responses become conditioned to stimuli as a function of their temporal contiguity to reinforcers. By contrast, Chomsky argued that the mind is sensitive to relations among interdependent elements of mental or linguistic representations that may be arbitrarily far apart (Fodor, 2008, p. 103).

The worry, then, was that the radical associationist could not account for why humans are so good at producing and understanding sentences that exhibit long-distance dependencies (especially if the relationships are very long).[ix] I will not discuss this objection in any detail here. What is important is that these supposed difficulties with radical associationism are often hailed as a chief advantage of the LOT hypothesis.

> On the one hand, there is no reason why computational relations need to be contiguity-sensitive, so if mental processes are computational they can be sensitive to `long-distance' dependencies between the mental (/linguistic) objects over which they are defined. On the other hand, we know that computations can be implemented by causal mechanisms; that's what computers do for a living. […] In short, a serious cognitive science requires a theory of mental processes that is compatible with the productivity of mental representations, and [the computational theory of mind] obliges by providing such a theory (Fodor, 2008, p.105).

Accounting for centre-embedded relative clauses may therefore be *crucial* for the LOT theorist who wishes to gain ground on the radical associationist (and Skinner). The LOT theorist will do this by explaining the comprehension of long-distance dependencies in English by positing recursive centre-embedded elements in mentalese sentences. If they do not take up this claim, then it is unclear how the relevant computations can be sensitive to long-distance dependencies. Theorists will, no doubt, be concerned about allowing English sentences to retain a structural complexity over and above that of mentalese sentences.[x]

I worry that if the LOT theorist posits recursive elements in mentalese, then she will also adopt another claim.

(2) Production and comprehension of centre embedded relative clauses is made possible by computational operations that exploit this syntactic feature of mental representations.

(2) Seems to be a natural consequence of (1) for the LOT theorist, because comprehending complex sentences will be a matter of composing complex mentalese expressions to represent the relevant propositions. But any psychological process for the computational theory of mind, including *understanding*, will be a matter of computational transformations of mental representations.

> If you think that a mental process - extensionally, as it were - as a sequence of mental states each specified with reference to its intentional content, then mental representations provide a mechanism for the construction of these sequences; and they allow you to get, in a mechanical way, from one such state to the next by *performing operations on the representations* (Fodor, 1987, p. 145).

Our theory of the composition of mental representations, then, constrains our theory of what kinds of computations can be performed on those mental representations. Any posited structural features had better be relevant to computation or there will be no reason for positing them in the first place. I will argue that modeling relative clause comprehension after the construction of complex mental representations (with centre-embedded recursive elements) places potentially problematic *constraints* on our theory of sentence processing.

3: Testing the LOT hypothesis

3.1: The problem with Recursive, Centre-Embedded, Elements

Some theorists have argued that the brain does not employ recursion in order to deal with centre-embedded relative clauses.

> It is an open question as to whether Chomsky's theories have any biologic validity. Can we really be sure that the relative, prepositional, and adverbial clauses that occur in complex sentences are derived from embedded sentences? […] In short, recursion as defined by Hauser, Chomsky, and Fitch (2002) follows from hypothetical syntactic structures, hypothetical insertions of sentences and phrase nodes, and hypothetical theory-specific rewriting operations on these insertions to form the sentences we actually hear or read (Lieberman, 2006, p. 359-360).

It may not be obvious that Lieberman's issue here is with a theory of neural processing (although one guesses this is what he means by "biologic validity"). However, Lieberman's argument is actually based on a theory of basal ganglia operations or, more specifically, a theory about the operations of segregated circuits that run from various areas of the cortex to segregated regions of the basal ganglia and back. I will postpone evaluating the evidential support for these claims until I explain their relevance to the LOT hypothesis. The argument begins from either of the following premises.

(I) Segregated cortical-subcortical circuits seem to perform similar computations (see, for example, Alexander & Crutcher, 1990; Delong, 1990) and these segregated circuits perform what can be considered motor, cognitive, and linguistic sequencing[xi] (Lieberman, 2006). One such circuit seems to be involved in the comprehension of centre-embedded relative clauses.

(II) A single circuit may perform cognitive and linguistic sequencing (see, for example, Lieberman, 2006; Hochstadt et al., 2006). This circuit is involved in the comprehension of centre-embedded relative clauses.

I propose that (I) and (II) are in tension with the claim that (1) the LOT permits recursive, centre-embedded, elements and the claim that (2) the comprehension of relative clauses is made possible by the recursive elements in LOT.

The basic problem posed by (I) is that motor control sequences are often thought to be less complex than linguistic ones. Devlin (2006) assumes, for instance, that motor sequences lack the hierarchical complexity of natural language. Not surprisingly, he predicts that the neural systems underlying complex behavioural sequences would employ fundamentally *different* kinds of computations, a proposition that is inconsistent with (I).

Comprehension of sentences with centre-embedded recursive elements, for the LOT theorist, should require a specific kind of recursion. "[…The] recursions […relevant to language] are defined over the constituent structure [and more specifically the *hierarchical constituent structure*] of mental representations" (Fodor, 2008, p. 105). If motor control sequences and linguistic sequences differ in complexity, then this difference needs to be played out in neural architecture.[xii]

(II) is problematic for similar reasons. If (II) is true, the LOT theorist will have to explain how the same computational system sequences cognitive and linguistic acts. The circuit in question is presumed to be important for the correct apprehension of certain centre embedded clauses. For example, patients with Parkinson's disease have difficulties comprehending a certain class of centre-embedded sentences from structural information alone (Hochstadt et al, 2006). Furthermore, the circuit also seems to be important for *cognitive set shifting*, a supposed component of executive function (Edabi & Pfeiffer, 2005).

If it turned out that a single system underlay the sequencing of both cognitive and linguistic acts, then this would burden the LOT theorist with making a unified account of the circuit's computations. If the LOT theorist accepts (1), then these computations had better be defined over the hierarchical-constituent-structure of these representations. LOT theorists must avoid, for instance, modeling the brain as performing *cognitive-shifts* at relative clauses boundaries (Hochstadt et al., 2006) unless these shifting operations are suitably defined over hierarchical-representations.

3.2: The Strength of the Case against LOT

I am not claiming that the current evidence is sufficient to overthrow the LOT hypothesis. There are, of course, plenty of presuppositions in Lieberman's argument that a theorist can object to. Note, however, that my case is made if (I) and (II) are respectable hypotheses that are in tension with the LOT hypothesis. I have only the burden of showing that (I) and (II) can be evaluated by serious research projects.

It is an old hypothesis that segregated cortical-subcortical-cortical circuits employ similar functions. Rather than viewing the basal ganglia as a single functional system taking projections from various cortical locations, evidence suggests that the basal ganglia are part of a number of largely (anatomically) segregated circuits (Alexander et al., 1986). The dorsolateral prefrontal cortex, for instance, seems to comprise a cortical-subcortical-circuit that is segregated from the motor circuit. Both circuits are composed of relatively segregated neural populations in the striatum, interior globus pallidus, and thalamus (Alexander et al., 1986).

It has also been suggested that this anatomical division might reflect a functional division (Alexander et al., 1986; Alexander & Crutcher, 1990). However, this functional division is to be made on the basis of those circuits' respective *cortical areas*. The circuits, themselves, seem to perform fundamentally similar operations.

> Because of the parallel nature of the basal ganglia-thalamocortical circuits and the apparent uniformity of synaptic organization at corresponding levels of these functionally segregated pathways […] it would seem likely that similar neuronal operations are performed at comparable stages of each of the five proposed circuits (Alexander et al., 1986, p. 361).

This poses a serious architectural hypothesis of cortical-subcortical-cortical circuits. What is still needed is a method for empirically testing (I) and (II). We must empirically ascertain whether these circuits have anything to do with language, motor control, or cognition. Two sources of evidence for this come from pathology and imaging data.

(i) Dysfunction in the Circuits (as seen in Parkinson's disease and Hypoxia) is associated with a wide array of cognitive, linguistic, and motor difficulties (including a difficulty with centre-embedded clauses) (Lieberman et al., 1990, 1992, 1995, 2005; Hochstadt et al, 2006; Lieberman, 2006). Furthermore, the PD linguistic difficulties may be due to a problem in comprehending relative clauses from structural information alone (Hochstadt, 2006).

(ii) Circuits are implicated with the performance of similar cognitive (for example, Monchi et al., 2001, 2004) and motor (for a brief review, see Edabi & Pfeiffer, 2005; Doyon et al., 2009) sequencing acts in healthy individuals.[xiii]

The most pressing obstacle to testing (i) and (ii) lies in the assumption that we can craft measures to track motor, linguistic, and cognitive sequencing.

A neural system's *sequencing powers* allow an organism to partake of certain complex behavioural sequences. These may be a complex motor sequence, such as that of jumping a hurdle while running, or that of producing a syntactically complex sentence. But even if we were satisfied with *motor* and *linguistic* sequencing, the notion of COGNITIVE SEQUENCING remains unclear. We are left to assume that what is meant (or perhaps should be meant) by *cognitive sequencing* is to be sorted out through the validation of our measures.

Lieberman views *the Wisconsin Card Sort* as measuring a capacity to *sequence* cognitive acts. This is crucial to his argument, as it ultimately leads him to view a cortical-subcortical-circuit as performing cognitive sequencing. The Wisconsin Card Sort is perhaps the most exhaustively studied measure that Lieberman relies upon, yet it is not obvious that the WCST is a valid measure of cognitive sequencing. It is, however, extremely common to view the WCST as measuring (among other things) a capacity for shifting one's strategic disposition to behave in response to changing environmental demands (see Flowers & Robertson, 1985, p. 517; Taylor & Saint-Cyr, 1995, p. 283; Edabi & Pfeiffer, 2005, p. 350). Set-shifting might be called upon, for example, when one must shift away from a failing chess strategy.

At least some of the measures Lieberman employs are in need of construct-validity. Hochstadt et al. (2006), for instance, employed the Test of Meaning from Syntax (TMS), which requires extensive further study before validity can be claimed. This test is crucial for his argument as it is presented here since it is taken to measure an individual's ability to comprehend centre-embedded relative clauses from syntactic information alone. Whether the TMS actually is such a measure will be contentious, regardless of its early promise (Hochstadt et al., 2006; Pickett, 1998). But if we believe in something like the Cronbach's & Meehl's (1955) account of *construct-validity*, then these difficulties are to be expected. Furthermore, whether (I) and (II) can be evaluated by serious research projects will turn on the respectability of construct-validity theory itself.

3.3: Construct-validity Theory

What makes construct-validity problematic is that it requires theorists to defend the claim that a test measures a *hypothetical construct*, a theoretical entity that cannot be directly observed (such as *aggression*). As a result, theorists need to draw upon a number of theoretical propositions in order to determine what kinds of observations should further construct-validity (Cronbach & Meehl, 1955). If a test battery really measures intelligence one might predict that it should predict academic success, though this requires the background assumption that academic performance is related to intelligence.

What seems to be the most contentious element of construct-validity is the idea that at the beginning of validation our theoretical constructs may be vague, but further study will make the meaning of our constructs clear. Cronbach & Meehl (1955) postulated a *nomological network* of scientific laws[xiv] in order to explain how demonstrating construct-validity could influence the meaning of a hypothetical construct. For Cronbach & Meehl, hypothetical constructs get their meaning from the role they play in this nomological network. "[…Even a] vague, avowedly incomplete network still gives the constructs whatever meaning they do have" (1955, p. 294). A nomological network might, for instance, have laws spelling out the consequences of *anger* on *aggression,* or the consequences of *heat* on *mercury expansion* in a thermometer. As we develop the network, and identify the consequences of a construct's satisfaction, our understanding of the construct is supposed to improve.[xv]

Lieberman needs such a view to support his belief that further empirical work will allow him to clarify what is meant by *cognitive sequencing*. More importantly, construct-validity theory allows theorists to use empirical methods to determine whether the TMS actually measures what it is supposed to. This will amount to defending a nomological network that explains *why* the TMS tracks the ability to comprehend centre-embedded relative clauses from structural information alone.

Assuming that some form of construct-validity theory holds,[xvi] theorists can continue the process of validating measures of motor, linguistic, and cognitive sequencing. Such measures will allow theorists to make claims about the importance of cortical-subcortical-circuits for various sequencing abilities. Results from such studies will put a burden on the LOT theorist, who hopes to account for complex sequencing behaviours in terms of computations on mental representations. At some point, these theorists must provide accounts of how similar computations will account for structurally complex motor, linguistic, and cognitive behaviour. Furthermore, these computations must make use of the composition of mental representations. If the LOT theorist posits centre-embedded recursive elements to account for the comprehension of relative clauses, then the computations they posit in explaining behaviour had better *exploit this composition*. If the comprehension of centre-embedded sentences (such as 'the boy who was fat kicked the clown' (Hochstadt et al., 2006)) has more in common with a cognitive set-shifting ability (as tracked by the Wisconsin Card Sort), then the LOT theorist must reconcile such facts with his theory of computation.[xvii]

This evidence may turn out to be a great boon for LOT theorists, as it may allow them to theorize about the composition of mental representations in a disciplined manner. However, it must be emphasized that language is a particularly important domain for the LOT hypothesis. The explanation of linguistic behaviour is supposed to be a task which, though an unsolvable mystery for radical associationism, is a manageable problem for a LOT (Fodor, 2008). But surely if the LOT hypothesis generates serious anomalies in the domain in which it is supposed to be most successful, then the entire hypothesis must come under suspicion.

References

Alexander, G. and M. Crutcher (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neuroscience*. *13*(7), 266–271.

Alexander, G. E., M. R. DeLong, and P. L. Strick (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience 9*, 357–381.

Austen, J. (1870). Northanger Abbey. London: Spottiswoode and Co.

Bedau, M. (1997). Weak emergence. In J. Tomberlin (Ed.), Philosophical Perspectives: Mind, Causation, and World., Volume 11, pp. 375–399. Malden, MA: Blackwell.

Carruthers, P. (2006). The Architecture of the Mind. Oxford University Press.

Chomsky, N. (1964). Current issues in linguistic theory. In J. A. Fodor and J. J. Katz (Eds.), The Structure of Language: Readings in the Philosophy of Language. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Chomsky, N. and G. A. Miller (1963). Introduction to the formal analysis of natural languages. In R. D. L.Galanter, R. R. Bush, and E. (Eds.), Handbook of Mathematical Psychology. New York and London: John Wiley and Sons, Inc.

Cronbach, L. J. and P. E. Meehl (1955). Construct validity in psychological tests. *Psychological Bulletin. 52*, 287–302.

de Saussure, F. (1959). Course in General Linguistics. New York: Philosophical Liberary.

Delong, M. (1990). Primate models of movement disorders of basal ganglia origin. *Trends in Neuroscience.13*(7), 281–285.

Devlin, J. (2006). Are we dancing apes? *Science. 314*(10).

Doyon, J., P. Bellec, R. Amsel, V. Penhune, O. Monchi, J. Carrier, S. Leh´ericy, and H. Benali. (2009). Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behavioral Brain Research. 199*(1), 61–75.

Edabi, M. S. and R. Pfeiffer (Eds.) (2005). Parkinson's Disease. CRC press.

Flowers, K. A. and C. Robertson (1985). The effect of Parkinson's disease on the ability to maintain a mental set. *Journal of Neurology, Neuroscience, and Psychiatry. 48*, 517–529.

Fodor, J. (1975). The Language of Thought. New York: Crowell.

Fodor, J. (1981). Introduction. In J. Fodor. (Ed.), Representations: Philosophical essays on the foundations of cognitive science. Hassocks: Harvester.

Fodor, J. (1987). Psychosemantics. Cambridge.: MIT press/Bradford Books.

Fodor, J. (2008). LOT 2: The Language of Thought Revisited. Oxford, England: Oxford University Press.

Gallistel, R. (1990). The Organization of Learning. MIT Press.

Gallistel, R. (2000). The replacement of general-purpose learning models with adaptively specialized learning modules. In M. Gazzaniga (Ed.), The New Cognitive Neurosciences, Second Edition. MIT Press.

Hauser, M. D., N. Chomsky, and W. T. Fitch (2002). The Faculty of Language, what is it, who has it, and how did it evolve? *Neuroscience 298*(22), 1569–1679.

Hochstadt, J., H. Nakano, P. Lieberman, and J. Friedman (2006). The roles of sequencing and verbal working memory in sentence comprehension deficits in Parkinson's disease. *Brain and Language 97*, 243–257.

Katz, J. and J. Fodor (1964). The structure of a semantic theory. In J. Katz and J. Fodor (Eds.), The Structure of Language: Readings in the Philosophy of Language. Englewood Cliffs, New Jersey.: Prentice-Hall, Inc.

Lieberman, P. (2006). Toward an Evolutionary Biology of Language. Cambridge Massachusetts:

The Belknap Press of Harvard University Press.

Lieberman, P., L. Feldman, and J. Friedman (1990). Syntax comprehension deficits in Parkinson's disease. *Journal of Nervous and Mental Disease 178*(360-365).

Lieberman, P., E. T. Kako, J. Friedman, G. Tajchman, L. S. Feldman, and E. B. Jiminez (1992). Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease. *Brain and Language 43*, 169–189.

Lieberman, P., B. G. Kanki, and A. Protopappas (1995). Speech production and cognitive decrements on Mount Everest. *Aviation, Space and Environmental Medicine 66*, 857–864.

Lieberman, P., A. Morey, J. Hochstadt, M. Larson, and S. Mather (2005). Mount Everest: A space-analogue for speech monitoring of cognitive deficits and stress. *Aviation, Space and environmental Medicine 76,* 198–207.

Monchi, O., M. Petrides, J. Doyon, R. Postuma, K. Worsley, and A. Dagher (2004). Neural bases of setshifting deficits in Parkinson's disease. *The Journal of Neuroscience. 24*(3), 702–710.

Monchi, O. and P. Petrides (2001). Wisconsin Card Sorting revisited: Distinct neural circuits participating in different stages of the task identified by event- related functional magnetic resonance imaging. *Neuroscience 21*, 7733–7741.

Pickett, E. (1998). Language and the Cerebellum. Ph. D. thesis, Brown University.

Putnam, H. (1973). Meaning and reference. *Journal of Philosophy. 70*, 669–711.

Rey, G. (1997). Contemporary Philosophy of Mind: A Contentiously Classical Approach. Blackwell Publishing.

Taylor, A. E. and J. A. Saint-Cyr (1995). The neuropsychology of Parkinson's Disease. *Brain and Cognition 28*, 281–296.

Wittgenstein, L. (1958). Philosophical Investigations. New York: The Macmillan Company.

Acknowledgements: This work owes a great deal to the suggestions of Ray Jennings and Kate Slaney. Thanks also goes to Trey Boone for suggesting that the evidence presented here might bear directly on the LOT hypothesis. This work was supported by SSHRC grant 766-2008-1100.

---

[i] What a mental representation actually *represents* is determined by (a) its syntactic composition and (b) the representational contents of its atomic parts[i] (Fodor, 2008). A representation's *causal powers* are exhausted by the

capacities of mutually supporting computational systems to track syntactic structure (computational systems do not track semantic content directly).

[ii] It should be emphasized that if conceptual attacks on the LOT hypothesis succeed, then the LOT hypothesis must be incoherent, and my work here will be largely irrelevant. I will assume, however, that LOT theorists will be unconvinced by such arguments. I only wish to argue, from within LOT's philosophical framework, that basal ganglia research poses empirical problems for the LOT hypothesis. Even if I am wrong, it may still turn out that the LOT hypothesis is incoherent.

[iii] Compositional mental representations allow for an intuitive kind of belief-desire psychology which I will call the computational-representational theory of mind (CRTM). The view implies that propositional attitudes are to be understood as relations between organisms and mental representations (Fodor, 2008; Rey, 1997). CRTM, however, takes this relationship to be spelled out in computational terms. An agent will have a belief that P if and only if they have a mental representation of P and this representation plays the right role in a computational system. This is supposed to provide an intuitive account of the causal efficacy of psychological properties. How my belief that P will influence my behaviour is to be determined by

(a) the fact that I *believe* that P rather than, say, desire that P

and

(b) the fact that I believe *that P* rather than *that Q*.

Manipulating the computational role of a representation (insofar as this is possible), and consequently manipulating an organism's relationship with a proposition, will have different behavioural consequences from those that involve manipulating the mental representation itself. This respects the intuition that a desire that P should have different behavioural consequences from those of a belief that Q.

[iv] The *syntactic* composition of a formal language is defined without appeal to semantics. Recursive rules generate a set of well formed sentences, and (because the class of atoms is well defined) their compositional structure can be defined with a notion of UNIFORM SUBSTITUTION. Semantic composition is defined *along* these recursive rules, ensuring that the semantic values of complex sentences are a function of the semantic values of the component parts.

Given independent definitions of a 'proof' and 'inference' we can seek to demonstrate the soundness and completeness of the proof theory with respect to the independent semantic theory of inference, and show that for every proof there is a corresponding valid inference, and vice versa. As a consequence, any machine that reliably generates proofs will also generate valid inferences, even though it has no direct access to our semantic interpretation of the language.

[v] There are a number of cases where systems are not supposed to be able to track what it is that they represent. A Putnam (1973) style example will suffice for these purposes. Imagine a possible world identical to ours, except that on twin-earth there is no H_20, only the compound *xyz*. H_20 and *xyz* have all the same surface properties, but a different underlying physical structure. Suppose that twin-earth-English contains the linguistic construction 'water', and (intuitively) this construction refers to *xyz*. By our previous supposition, human neural systems are incapable of distinguishing H_20 from *xyz*. The twin-earth mentalese expression representing H_20 could be syntactically identical to the earth representation of. A computational system will not be sensitive to whether it represents H_20 or *xyz*.

[vi] It should be noted that, on the present understanding of the LOT hypothesis, sentences in mentalese need not be the contents of *thought*. Thoughts, beliefs, desires, and other psychological states, according to the computational theory of mind, imply certain computational relationships between an organism and a representation. A mentalese sentence, on its own, is merely a compositionally structured mental representation.

[vii] The triangles stand in for a collection of lower branches.

[viii] I distinguish the claim that some part of a mental representation represents a noun phrase and the claim that noun phrases are part of the structure of mental representations. The latter is a claim about the syntax of LOT, while the former is a claim about semantics.

[ix] It is important to note here that Fodor (2008) may not have the relatively short dependencies of Jane Austen in mind. The productivity of language arguably extends into the domain of long-distance dependencies, and it sometimes seems as if English grammar allows for an arbitrary number of centre-embeddings.

The boy, who lifted the ball, kicked the dog.

The boy, who lifted the ball, which was lighter than 1 kilogram, kicked the dog.

The boy, who lifted the ball, which was lighter than one kilogram, which is lighter than 2 kilograms, kicked the dog.

and so on…

Our linguistic intuitions are supposed to suggest that the sentences picked out here are all grammatical, along with sentences containing any finite number of embeddings.

[x] For a fuller discussion of the importance of centre-embedded recursive elements to sequence complexity, see Chomsky (1963)

[xi] For a discussion of what is meant by *sequencing* see 3.2.

[xii] One way to avoid this problem is to accept, as Lieberman (2006) does, that motor sequences are hierarchically complex. But even this does not completely dissolve the difficulty. Eventually, the LOT theorist will have to propose similar computational processes for the two circuits. These computations will have to be defined over centre-embedded recursive elements and account for both linguistic and motor sequencing.

[xiii] I know of no strong evidence for the claim that cortical-subcortical-cortical circuits are involved in the comprehension of centre-embedded relative clauses *in healthy individuals*. Lieberman (personal communication) does predict, however, that imaging studies should find basal ganglia activation during this kind of linguistic sequencing.

[xiv] Cronbach & Meehl (1955) made no assumptions regarding whether the laws were statistical or deterministic.

[xv] Theorists may want to distinguish *possession of a concept* and *explication of a concept* in order to fit construct-validity theory into contemporary semantic theories. If an individual possesses a concept then they will have a mental representation corresponding to the concept. Possessing the concept DOG is necessary for Jennifer to think 'that is a dog'. Possession of a concept does not necessarily imply, however, that one can explicate a concept. When asked what dogs are, there is no guarantee that Jennifer can say anything useful. One might, then, view nomological networks as influencing only our ability to *explicate* concepts, and not our capacity to posses them.

[xvi] Some may view construct-validity theory as objectionable, but this will not be true for LOT theorists. The cognitive psychologist needs some way of measuring *mental processes*. A theorist cannot simply stare at a brain until a computational theory dawns. Theorists need to *develop measures* that will track underlying computational

processes. They will have no idea what kind of computational relationship will correspond to a *belief* until the science is well under way. They must rely upon some form of construct-validity theory to justify a methodology of starting from a loose understanding of beliefs as *some* kind of computational relationship between an organism and its mental representations. In other words, construct-validity theory relieves the scientist of the burden of laying down a computational theory of *belief* in advance. As theorists fill in the details of the nomological network (and develop a computational theory of mind) they are supposed to come closer to explicating the concept of BELIEF.

[xvii] Reconciling the LOT hypothesis with these facts may still be possible. Theorists can try to explain cognitive set-shifting with computations that are defined over recursive elements. Perhaps *cognitive sets*, themselves, contain recursive elements. Theorists are welcome to respond this way to anomalous results, but their responses will not be insulated from revision.