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.

> 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

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]

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

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.

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

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.

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.

Cronbach, L. J. and P. E. Meehl (1955). Construct valiEmd–3 1EmCe

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.

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: !"#$%&'()%'&*$%+%,(*+-%.*+/%-'%-"*%$0,,*$-#'1$%'2%3+4%5*11#1,$%+1.%6+-*%7/+1*48% !"+1)$%+/$'%,'*$%-'%!(*4%9''1*%2'(%$0,,*$-#1,%-"+-%-"*%*:#.*1;*%<(*$*1-*.%"*(*%=#,"-%>*+(%.#(*;-/4%'1% -"*%?@!%"4<'-"*$#$8%!"#$%&'()%&+$%$0<<'(-*.%>4%77A3B%,(+1-%CDDEFGGHEIIGG8

[#]%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).

[##] 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.

[###] 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.

[#:] 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

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-