The computational and the representational language of thought hypotheses

David J. Chalmers

New York University

chalmers@nyu.edu, consc.net/chalmers

Behavioral and Brain Sciences, forthcoming. Commentary on Jake Quilty-Dunn, Nicolas Porot, and Eric Mandelbaum, "The Best Game in Town: The Re-Emergence of the Language of Thought Hypothesis Across the Cognitive Sciences", *Behavioral and Brain Sciences,* 2022.

Abstract: There are two versions of the language of thought hypothesis (LOT): representational LOT (roughly, structured representation), introduced by Ockham, and computational LOT (roughly, symbolic computation) introduced by Fodor. Like many others, I oppose the latter but not the former. Quilty-Dunn, Porot, and Mandelbaum defend representational LOT, but they do not defend the strong computational LOT thesis central to the classical-connectionist debate.

There are two versions of the language of thought hypothesis. The representational language of thought hypothesis (r-LOT), introduced by William of Ockham and defended by Quilty-Dunn, Porot, and Mandelbaum (QPM), concerns the structure of mental representation. The computational language of thought hypothesis (c-LOT), introduced by Jerry Fodor, concerns computation over mental representations. Representational LOT is much weaker than computational LOT and is more widely accepted. I accept the former but reject the latter. As a result, I agree with many of QPM's conclusions while finding that they have not really defended the most controversial form of LOT.

In more detail: r-LOT (I use "LOT" for both the language and the hypothesis) says roughly that thought involves sententially structured mental representations. At minimum, there are nominal representations (e.g. *Biden*) and predicative representations (e.g. *president*) that combine into structured representations (e.g. *Biden is president*) with propositional content. Structured representations may also involve connectives (e.g. *and*), quantifiers (e.g. *all*), operators (e.g. *always*), and other types familiar from the linguistic case.

R-LOT is not trivially true, but it is plausible and hard to deny. It follows naturally from the claims that (1) people make judgments such as *Biden is president*, (2) these judgments involve combining nominal and predicative representations (or concepts, in the sense where concepts are mental representations) such as *Biden* amd *president*, and (3) these representations can be recombined in judgments such as *Biden is from Delaware*. My sense is that most contemporary cognitive scientists and philosophers of mind accept these fairly weak claims. Importantly, these claims do not have immediate consequences regarding computation or cognitive architecture.

The computational language of thought hypothesis (c-LOT) adds to r-LOT the key claim that thought involves computation over these sententially structured representations. The classical version of this hypothesis says that r-LOT representations are the medium through which all cognitive computation takes place. That is, the basic vehicles of representation in the r-LOT system (atomic words in the representational language of thought) also serve as the basic vehicles of computation (atomic computational states to which cognitive algorithms apply).

The computational language of thought hypothesis was canonically formulated in Fodor's *The Language of Thought* (1975). Computation plays a central role throughout the book, from the main argument for LOT at the start of chapter 1 ("Computation presupposes a medium of computation: a representational system", p. 27) to the conclusion ("More exactly: Mental states are relations between organisms and internal representations, and causally interrelated mental states succeed one another according to computational principles which apply formally to the representations", p. 198). There are other works (e.g. "Propositional attitudes") in which Fodor focuses mainly on r-LOT, but computation is central in the canonical statement.

(Related distinctions: Fodor himself (1980) distinguishes the "representational theory of mind" and the "computational theory of mind" (though neither requires a language of thought). Michael Rescorla (2017) distinguishes a core version of LOT that involves "representational theory of thought" plus "compositionality of thought" and perhaps "logical structure" (in my terms, a version of r-LOT) from a stronger version that adds "the classical computational theory of mind" (yielding a version of c-LOT, though I understand the computational constraint differently from Rescorla).)

Most work in symbolic AI uses a version of c-LOT. Both involve computation over atomic symbols: entities that are both representationally atomic and computationally atomic. Atomic symbols have no computationally relevant internal structure (if they did, they would not be computationally atomic). Instead, their internal form is arbitrary.

The most significant opposition to LOT, in the classical-connectionist debate, has been opposition to c-LOT. In most neural network models there are no computationally atomic symbols. Representations are distributed over multiple quasi-neural units. As a result, in these models computation is *subsymbolic computation*: computation takes place among units below the level of representation. Because computational primitives (units) are not representational primitives in these models, representation is not the medium of computation. Subsymbolic computation is incompatible wth c-LOT.

At the same time, subsymbolic computation is quite compatible with r-LOT. This is clearest in the work of structured connectionists (e.g. Smolensky 1998, Chalmers 1990), where distributed representations (e.g. of *Biden* and *president*) can combine with each other systematically to yield new distributed representations such as *Biden is president*. This is naturally seen as a structured representational system involving subsymbolic computation: r-LOT without c-LOT. The structured connectionist research program is still a work in progress, but it is arguable that contemporary large language models also combine structured representation (of facts such as *Biden is president*) with subsymbolic computation. A second and third way of combining r-LOT with subsymbolic computation are provided by the framework of vector symbolic architectures (Kleyko et al 2022), where representations are vectors, and Piantadosi's combinator framework (2021), where the computational primitives S and K fall below the level of representation.

(Terminology: all three of these are computational versions of r-LOT in a broad sense. In an alternative phraseology, one might call the Fodorian version the *classical* computational LOT (cc-LOT), while calling subsymbolic versions *nonclassical* computational LOT (nc-LOT). But I will reserve "c-LOT" for the classical Fodorian version.)

Proponents of LOT often argue that structured connectionism is merely an implementation of LOT. We can now see that this claim is false or at best misleading. Implementation is standardly a computational relation between algorithms, requiring the implementing algorithm to be a more fine-grained version of the implemented algorithm with the same input/output behavior. The most interesting subsymbolic algorithms (e.g. in artificial neural networks) are never implementations of symbolic algorithms in this sense. The success of the deep learning paradigm has provided strong evidence that the behavior of these systems (especially their success in learning and generalizing, but also their post-learning success) is not the result of implementing a more coarse-grained symbolic algorithm and cannot be duplicated by such algorithms. These systems may realize an r-LOT, but they do not implement a c-LOT. The quasi-symbolic operations of composition, decomposition, and quasi-logical inference may be available, but they are a tiny subset of the operations one can perform on the relevant distributed representations. As I argued in Chalmers (1990), one can also perform all sorts of holistic operations on distributed representations that do not proceed via these symbolic operations. It is plausibly subsymbolic operations like this that are largely responsible for the remarkable capacities of neural network systems.

QPM don't make the distinction between r-LOT and c-LOT in their article, but their LOT appears to be a version of r-LOT. Their six core claims defining LOT do not mention computation (except in one case, incidentally). Four of the key claims (role-filler independence, predicate-argument structure, logical operators, abstract conceptual content) clearly pertain to representation but not computation. A fifth (inferential promiscuity) mentions computational theories of logical inference as versions of LOT, but computation does not play a defining role, and inferential promiscuity can equally be present in r-LOT without c-LOT (e.g. Ockham-style or subsymbolic systems).

The requirement of "discrete constituents" may suggest c-LOT, though it doesn't mention computation explicitly. Distributed representations in a structured connectionist systems arguably aren't discrete in the authors sense, in that representation of *Biden* and of *president* (say) can be intertwined non-discretely in a representation of *Biden is president*. On the other hand, many subsymbolic computational systems involve discrete constituents without c-LOT. Piantadosi's system is one. Another is provided by the word embedding format for representing words that is ubiquitous in current language models. Here words are represented by multidimensional vectors where individual units often lack any clear semantic significance. "Biden is president" may be represented as a sequence of vectors for the individual words, so the constituents are discrete, but representations remain distributed and processing remains subsymbolic. So the discrete representational constituents does not require c-LOT.

Now, perhaps the absence of a computational constraint is an easily correctable omission. QPM discuss computational approaches at some length in other sections of their article. They could easily enough add a seventh constraint connecting computation to representation, holding that the representational primitives are computationally primitive and serve as the medium of computation. The trouble is that strong evidence for this seventh claim is much harder to find.

The target article does argue that many Bayesian theorists provide computational accounts involving a "probabilistic LOT" associated with sententially structured representations. This suggests r-LOT, but it does not obviously lead to c-LOT, as Bayesian accounts are usually not cast at the algorithmic level (rather, at Marr's higher "computational" level). These accounts have many algorithmic implementations, including subsymbolic implementations in deep learning systems. So there is no obvious strong evidence for c-LOT here, and any evidence would need to be stacked against the counterevidence provided by deep learning models.

Overall: if QPM are defending c-LOT, then more work is needed to make the defense explicit. If they are defending only r-LOT, then their conclusion is plausible, and my only objection is one of relative unambition.

Bibliography

Chalmers, D. J. (1990). Syntactic Transformations on Distributed Representations. Connection Science, 2(1), 53-62.

Fodor, J. A. (1975). The Language of Thought. Harvard University Press.

Fodor, J. A. (1980). Representations: Philosophical Essays on the Foundations of Cognitive Science. MIT Press.

Kleyko, D., Davies, M., Frady, E. P., Kanerva, P., Kent, S. J., Olshausen, B. A., Osipov, E., & Rabaey, J. M. (2022). Vector Symbolic Architectures as a Computing Framework for Emerging Hardware. *Proceedings of the IEEE.* 110:1538.

Piantadosi, S. T. (2021). The computational origin of representation. *Minds and Machines* 31:1-58.

Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2023). The Best Game in Town: The Re-Emergence of the Language of Thought Hypothesis Across the Cognitive Sciences. *Behavioral and Brain Sciences*.

Rescorla, M. (2017). The Language of Thought Hypothesis. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Smolensky, P. (1988). On the Proper Treatment of Connectionism. *Behavioral and Brain Sciences*, 11(1), 1-23.