The Language of Thought Hypothesis

The language of thought hypothesis (LOTH) is the hypothesis that mental representation has a linguistic structure, or in other words, that thought takes place within a mental language. The hypothesis is sometimes expressed as the claim that thoughts are sentences in the head. It is one of a cluster of other hypotheses that together offer a theory of the nature of thought and thinking. The other hypotheses in the cluster include the causal-syntactic theory of mental processes (CSMP), and the representational theory of mind (RTM). The former is the hypothesis that mental processes are causal processes defined over the syntax of mental representations. The latter is the hypothesis that propositional attitudes are relations between subjects and mental representations. Taken together these theses purport to explain how rational thought and behavior can be produced by a physical object, such as the human brain. In short, the explanation is that the brain is a computer and that thinking is a computational process. The cluster therefore is referred to often (and aptly) as the computational theory of mind (CTM).

LOTH was first introduced by Jerry Fodor in his 1975 book *The Language of Thought*, and further elaborated and defended in a series of works by Fodor and several collaborators. Fodor's original argument for LOTH rested on the claim that (at the time) the only plausible psychological models presupposed linguistically structured mental representations. Subsequent arguments for LOTH are inferences to the best explanation. They appeal to supposed features of human cognition such as productivity, systematicity, and inferential coherence, arguing that these features are best explained if LOTH is true. Important objections to LOTH have come from those who believe that the mind is best modeled by connectionist networks, and by those who believe that (at least some) mental representation takes place in other formats, such as maps and images.

This article has three main sections. The first explains LOTH, as well as CSMP, RTM, and the importance of conjoining all three to arrive at the resulting CTM. The second describes the major arguments in favor of LOTH. The third describes some important problems for LOTH and objections to it.

Table of Contents

1. The Language of Thought Hypothesis
   a. Combinatorial Syntax and Compositional Semantics
   b. Mental Processes as Causal-Syntactic Processes
   c. RTM and the Propositional Attitudes
   d. The Computational Theory of Mind
   e. Theories of Meaning
2. Arguments for LOTH
   a. The Only Game in Town
   b. Productivity
   c. Systematicity
   d. Inferential Coherence
3. Problems and Objections
   a. Individuating Symbols
   b. Context-dependent Properties of Thought
   c. Mental Images
   d. Mental Maps
   e. Connectionist Networks
   f. Analog and Digital Representation
4. References and Further Reading

1. The Language of Thought Hypothesis

   a. Combinatorial Syntax and Compositional Semantics

    LOTH is the claim that mental representation has a linguistic structure. A representational system has a linguistic structure if it employs both a *combinatorial syntax* and a *compositional semantics* (see Fodor and Pylyshyn 1988 for this account of linguistic structuring).

    A representational system possesses a combinatorial syntax if,
    
    (i) it employs two sorts of representation: atomic representations and compound representations, and
    (ii) the constituents of compound representations are either compound or atomic.

    A representational system possesses a compositional semantics if,

    (iii) the semantic content of a representation is a function of the semantic content of its syntactic constituents, the overall structure of the representation, and the arrangement of the constituents within the overall structure.
Formal languages are good examples of languages possessing both combinatorial syntax and compositional semantics. For example, sentential logic (propositional logic) employs symbols to represent simple declarative sentences (usually the capital letters ‘A’, ‘B’, ‘C’,…) and symbols for logical connectives (usually ‘¬’ for ‘and’, ‘∨’ for ‘or’, ‘→’ for ‘if…then.…’, and so on). Thus, ‘A’ might be an atomic representation of the sentence ‘Gail is tall’, ‘B’ an atomic representation of the sentence ‘Alan is bald’, and ‘C’ an atomic representation of the sentence ‘Amanda is funny’. In that case, (A · B) v C would be a compound representation of the sentence ‘Either Gail is tall and Alan is bald, or Amanda is funny’. The components of this compound representation are the compound representation (A · B) and the atomic representation ‘C’. In short, sentential logic employs both atomic and compound representations, and the components of its compound representations are themselves either atomic or compound. Thus, it possesses a combinatorial syntax. Moreover, the semantic content of a representation within sentential logic (generally taken to be a truth-value—either TRUE or FALSE) is a function of the content of the syntactic constituents, together with overall structure and arrangement of the representation. For instance, the truth-value of a representation with the form ‘A → B’ is TRUE just in case the truth-value of ‘A’ is FALSE or the truth-value of ‘B’ is TRUE. Alter the arrangement of the parts (B → A) or the overall structure (A · B) or the components (A → C) and the truth-value of the whole may change as well. Therefore it also possesses a compositional semantics.

LOTH amounts to the idea that mental representation has both a combinatorial syntax and a compositional semantics. It is the idea that thoughts occur in a formal mental language termed the “language of thought” or often “mentalese”). A common way of casting it is as the claim that thoughts are literally sentences in the head. This way of explaining the thesis can be both helpful and misleading.

First, it is important to note that sentences can be implemented in a multitude of different kinds of media, and they can be written in a natural language or encoded in some symbolic language. For example, they may be written on paper, etched in stone, or encoded in the various positions of a series of electrical switches. They may be written in English, French, first-order logic, or Morse code. LOTH claims that at a high level of abstraction, the brain can be accurately described as encoding the sentences of a formal language.

Second, it is equally important to note that the symbolic language LOTH posits is not equivalent to any particular spoken language but is the common linguistic structure in all human thought. Part of Fodor’s (1975) original argument for LOTH was that learning a spoken language requires already possessing an internal mental language, the latter being common to all members of the species.

Third, the posited language is not appropriately thought of as being introspectively accessible to a thinking subject. In other words, while thinkers may have access to much of what goes on while they are thinking (for example the images, words and so on that may be visible “in the mind’s eye”), the language of thought is not “visible” as such. Rather, it is best thought of as the representations that are being tokened in and processed by the brain, during and “beneath” all that is accessible to the thinker. (However, that they are not introspectively accessible is not to be taken to indicate that they are not causally efficacious in the production of behavior. On the contrary, they must be, if the theory is to explain the production of rational behavior.)

Casting LOTH as the idea of sentences in the head can be useful, if understood appropriately: as sentences of a species-wide formal language, encoded in the operations of the brain, which are not accessible to the thinker.

b. Mental Processes as Causal-Syntactic Processes

Representational systems with combinatorial syntax and compositional semantics are incredibly important, as they allow for processes to be defined over the syntax of the system of representations that will nevertheless respect constraints on the semantics of those representations. For example, standard rules of inference for sentential logic—rules such as modus ponens, which allows the inference from a representation of the form ‘A → B’ together with a representation of the form ‘A’ to a representation of the form ‘B’—are defined over the syntax of the representations. Nevertheless, the rules respect the following semantic constraint: given true premises, correct application of them will result only in true conclusions.

Processes defined over the syntax of representations, moreover, can be implemented in physical systems as causal processes. Hence, representational systems possessing both combinatorial syntax and compositional semantics allow for the construction of physical systems that behave in ways that respect the semantic constraints of the implemented representational system. That is, they allow for the construction of machines that “think” rationally. Modern digital computers are just such machines: they employ linguistically structured representations and processes defined over the syntax of those representations, implemented as causal processes.

Since LOTH is the claim that mental representation has both combinatorial syntax and compositional semantics, it allows for the further claim that mental processes are causal processes defined over the syntax of mental representations, in ways that respect semantic constraints on those representations (Fodor 1975, Fodor and Pylyshyn 1988). This further claim is the causal-syntactic theory of mental processes (CSMP). LOTH and CSMP together assert that the brain, like a digital computer, processes linguistically structured representations in ways that are sensitive to the syntax of those representations. Indeed, the advent of the digital computer inspired CTM. This will be further discussed below.

c. RTM and the Propositional Attitudes

LOTH is a specification of the representational theory of mind (RTM). RTM is the thesis that commonsense mental states, the propositional attitudes such as believing, desiring, hoping, wishing, and fearing are relations between a subject and a mental representation. According to RTM, a propositional attitude inherits its content from the content of the representation to which the thinker is related. For example, Angie believes that David stole a candy bar if and only if there is a belief relation between Angie and a mental representation, the content of which is David stole a candy bar. Thus, where ‘φ’ names a propositional attitude, and ‘p’ is the content of a propositional attitude, a technical rendering of RTM is as follows:

(R1) A subject S φ’s that p if and only if there is a relation Rφ and a mental representation P such that S bears Rφ to P and P means that p.

According to RTM, the difference between Angie’s believing that David stole a candy bar and her hoping that David stole a candy bar, lies in there being different relations between her and the same representation of the content David stole a candy bar. Thus, (R1) is a schema. For specific propositional attitudes, the name of the attitude will take the place of ‘φ’ in the schema. For example, the case of belief is as follows:
RTM is a species of intentional realism—the view that propositional attitudes are real states of organisms, and in particular that a mature psychology will make reference to such states in the explanation of behavior. For debate on this issue see for example Churchland 1981, Stich 1983, Dennett 1987. One important virtue of RTM is that it provides an account of the difference between the truth and falsehood of a propositional attitude (in particular, of a belief). On that account, the truth or falsehood of a belief is inherited from the truth or falsehood of the representation involved. If the relationship of belief holds between Angie and a representation with the content David stole a candy bar, yet David did not steal a candy bar, then Angie has a false belief. This account also provides an explanation of the so-called “Fregian cases” in which a subject believes that a given object known by one name has some property yet the subject fails to believe that the same object known by another name has the same property (see Fodor 1978).

d. The Computational Theory of Mind

RTM, LOTH, and CSMP was inspired on one hand by the development of modern logic, and in particular by the formalization of logical inference (that is, the development of rules of inference that are sensitive to syntax but that respect semantic constraints). On the other hand, it was inspired by Alan Turing's work showing that formal procedures can be mechanized, and thus, implemented as causal processes in physical machines. These two developments led to the creation of the modern digital computer, and Turing (1950) argued that if the conversational behavior (via teletype) of such a machine was indistinguishable from that of a human being, then that machine would be a thinking machine. The combination of RTM, LOTH, and CSMP is in a sense the converse of this latter claim. It is the idea that the mind is a computer, and that thinking is a computational process. Hence the combination of these theses has come to be known as the Computational Theory of Mind (CTM).

The importance of CTM is twofold. First, the idea that thinking is a computational process involving linguistically structured representations is of fundamental importance to cognitive science. It is among the origins of work in artificial intelligence, and though there has since been much debate about whether the digital computer is the best model for the brain (see below) many researchers still presume linguistic representation to be a central component of thought.

Second, CTM offers an account of how a physical object (in particular, the brain) can produce rational thought and behavior. The answer is that it can do so by implementing rational processes as causal processes. This answer provides a response to what some philosophers—most famously Descartes, have believed: that explaining human rationality demands positing a form of existence beyond the physical. That is, it is a response to dualism (see Descartes 1637/1985, 139-40, and see Rey 1997 for discussion of CTM as being a solution to "Descarte's challenge"). It therefore stands as a major development in the philosophy of mind.

c. Theories of Meaning

Explaining rationality in purely physical terms is one task for a naturalized theory of mind. Explaining intentionality (the meaning or "aboutness" of mental representations) in purely physical terms is a related, though separate, task for a naturalized theory of mind. Famously, Brentano (1874/1995) worried that intentionality cannot be explained in physical terms, as Descartes believed rationality could not be explained in physical terms (see Rey 1997 for CTM being a solution to "Brentano's challenge").

Still, CTM lends itself to a physicalist account of intentionality. There are two general strategies here. Internalist accounts explain meaning without making mention of any objects or features external to the subject. For example, conceptual role theories (see for instance Loar 1981) explain the meaning of a mental representation in terms of the relations it bears to other representations in the system. Externalist accounts explicitly tie the meaning of mental representations to the environment of the thinker. For example, causal theories (see for instance Dretske 1981) explain meaning in terms of causal regularities between environmental features and mental representations.

Fodor (1987) has argued for an “asymmetric dependency theory,” which is a kind of causal theory of meaning, intended specifically to deal with the disjunction problem that plagues causal theories. The problem arises for causal theories of meaning because of the seemingly obvious fact that some mental representations are caused by objects they do not represent. For example, on a dark evening, someone might easily mistake a cow for a horse; in other words, a cow might cause the tokening of a mental representation that means horse. But if, as causal theories have it, the meaning of a representation is determined by the object or objects that cause it, then the meaning of such a representation is not horse, but rather horse or cow (since the type of representation is sometimes caused by horses and sometimes caused by cows).

Fodor's solution is to suggest that such a representation means horse and not horse or cow, because the fact that it may sometimes be caused by cows is dependent on the fact that it is usually caused by horses. That is, if the representation was not caused by horses, then it would not sometimes be caused by cows. But this dependence is asymmetric: if the representation was not ever caused by horses, it would nevertheless still be caused by horses. CTM, and LOTH in particular, need not be wedded to Fodor’s account. As all of the above examples explain meaning in physical terms, the coupling of a successful CTM with a successful version of any of them would yield an entirely physical account of two of the most important general features of the mind: rationality and intentionality.

2. Arguments for LOTH

LOTH then, is the claim that mental representations possess combinatorial syntax and compositional semantics—that is, that mental representations are sentences in a mental language. This section describes four central arguments for LOTH. Fodor (1975) argued that LOTH was presupposed by all plausible psychological models. Fodor and Pylyshyn (1988) argue that thinking has the properties of productivity, systematicity, and inferential coherence, and that the best explanation for such properties is a linguistically structured representational system.

a. The Only Game in Town

Fodor’s (1975) argument for LOTH proceeded from the claim that the only “remotely plausible” models of cognition are computational models. Because computational models presuppose a medium of representation, in particular a linguistic medium, and because “remotely plausible theories are better than no theories at all,” Fodor claimed that we were “provisionally committed” to LOTH. In short, the
argument was that the only game in town for explaining rational behavior presupposed internal representations with a linguistic structure.

The development of connectionist networks—computational systems that do not presuppose representations with a linguistic format—therefore pose a serious challenge to this argument. In the 1980s, the idea that intelligent behavior could be explained by appeal to connectionist networks grew in popularity and Fodor and Pylyshyn (1988) argued on empirical grounds that such an explanation could not work, and thus that even linguistic computation was no longer the only game in town, it was still the only plausible explanation of rational behavior. Their argument rested on claiming that thought is productive, systematic, and inferentially coherent.

b. Productivity

Productivity is the property a system of representations has if it is capable, in principle, of producing an infinite number of distinct representations. For example, sentential logic typically allows an infinite number of sentence letters (A, B, C, ...), each of which is a unique atomic representation. Thus the system is productive. A street light, on the other hand, has three atomic representations (“red”, “yellow”, “green”), and no more. The system is not productive. Productivity can be achieved in systems with a finite number of atomic representations, so long as those representations may be combined to form compound representations, with no limit on the length of the compounds. Here are three examples: A, A → B and ((A → B) → B). That is, productivity can be achieved with finite means by employing both combinatorial syntax and compositional semantics.

Fodor and Pylyshyn (1988) argue that mental representation is productive, and that the best explanation for its being so is that it is couched in a system possessing combinatorial syntax and compositional semantics. They first claim that natural languages are productive. For example, English possesses only a finite number of words, but because there is no upper bound on the length of sentences, there is no upper bound on the number of unique sentences that can be formed. More specifically, they argue that the capacity for sentence construction of a competent speaker is productive—that is, competent speakers are able to create an infinite number of unique sentences. Of course, this is an issue in principle. No individual speaker will ever construct more than a finite number of unique sentences. Nevertheless, Fodor and Pylyshyn argue that this limitation is a result of having finite resources (such as time).

The argument proceeds by noting that, just as competent speakers of a language can compose an infinite number of unique sentences, they can also understand an infinite number of unique sentences. Fodor and Pylyshyn write,

> there are indefinitely many propositions which the system can encode. However, this unbounded expressive power must presumably be achieved by finite means. The way to do this is to treat the system of representations as consisting of expressions belonging to a generated set. More precisely, the correspondence between a representation and the proposition it expresses is, in arbitrarily many cases, built up recursively out of correspondences between parts of the expression and parts of the proposition. But, of course, this is not possible. No individual speaker will ever construct more than a finite number of unique sentences. Nevertheless, Fodor and Pylyshyn argue that this limitation is a result of having finite resources (such as time).

In short, human beings can entertain an infinite number of unique thoughts. But since humans are finite creatures, they cannot possess an infinite number of unique atomic mental representations. Thus, they must possess a system that allows for construction of an infinite number of thoughts given only finite atomic parts. The only systems that can do that are systems that possess combinatorial syntax and compositional semantics. Thus, the system of mental representation must possess those features.

c. Systematicity

Systematicity is the property a representational system has when the ability of the system to express certain propositions is intrinsically related to the ability the system has to express certain other propositions (where the ability to express a proposition is just the ability to token a representation whose content is that proposition). For example, sentential logic is systematic with respect to the propositions Bill is boring and Fred is funny and Fred is funny and Bill is boring, as it can express the former if and only if it can also express the latter. Similarly to the argument from productivity, Fodor and Pylyshyn (1988) argue that thought is largely systematic, and that the best explanation for its being so is that mental representation possesses a combinatorial syntax and compositional semantics.

The argument rests on the claim that the only thing that can account for two propositions being systematically related within a representational system is if the expressions of those propositions within the system are compound representations having the same overall structure and the same components, differing only in the arrangement of the parts within the structure, and whose content is determined by structure, parts, and arrangement of parts within the structure. Thus, the reason the propositions Bill is boring and Fred is funny and Fred is funny and Bill is boring are systematically related in sentential logic is because the representation of the former is ‘B · F’ and the representation of the latter is ‘F · B’. That is, they are both conjunctions, they have the same components, they only differ in the arrangement of the components within the structure, and the content of each is determined by their structure, their parts, and the arrangement of the parts within the structure. But, the argument continues, any representational system that possesses multiple compound representations that are capable of having the same constituent parts and whose content is determined by their structure, parts and arrangement of parts within the structure is a system with combinatorial syntax and compositional semantics. Hence, systematicity guarantees linguistically structured representations.

Fodor and Pylyshyn argue that, if thought is largely systematic, then it must be linguistically structured. They argue that for the most part it is, pointing out that anyone who can entertain the proposition that John loves Mary can also entertain the proposition that Mary loves John. What explains that is that the underlying representations are compound, have the same parts, and have contents that are determined by the parts and the arrangement of the parts within the structure. But then what underlies the ability to entertain those propositions is a representational system that is linguistically structured. (See Johnson 2004 for an argument that language, and probably thought as well, is not systematic).

d. Inferential Coherence

A system is inferentially coherent with respect to a certain kind of logical inference, if given that it can draw one or more specific inferences that are instances of that kind, it can draw any specific inferences that are of that kind. For example, let A be the proposition Emily is in Scranton and Judy is in New York, and let B be the proposition Emily is in Scranton. Here A is a logical conjunction, and B is the first conjunct. A system that can draw the inference from A to B is a system that is able to infer the first conjunct from a conjunction with two conjuncts, in at least one instance. A system may or may not be able to do the same given other instances of the same kind of inference. It
may not for example be able to infer Bill is boring from Bill is boring and Fred is funny. If it can infer the first conjunct from a logical conjunction regardless of the content of the proposition, then it is inferentially coherent with respect to that kind of inference. As with productivity and systematicity, Fodor and Pylyshyn point to inferential coherence as a feature of thought that is best explained on the hypothesis that mental representation is linguistically structured.

The argument here is that what best explains inferential coherence with respect to a particular kind of inference, is if the syntactic structure of the representations involved mirrors the semantic structure of the propositions represented. For example, if all logical conjunctions are represented by syntactic conjunctions, and if the system is able to separate the first conjunct from such representations, then it will be able to infer for example, Emily is in Scranton from Emily is in Scranton and Judy is in New York, and it will also be able to infer Bill is boring from Bill is boring and Fred is funny, and so on for any logical conjunction. Thus it will be inferentially coherent with respect to that kind of inference. If the syntactic structure of all the representations matches the logical structure of the propositions represented, and if the system has general rules for processing those representations, then it will be inferentially coherent with respect to any of the kinds of inferences it can perform.

Representations whose syntactic structure mirrors the logical structure of the propositions they represent, however, are representations with combinatorial syntax and compositional semantics; they are linguistically structured representations. Thus, if thought is inferentially coherent, then mental representation is linguistically structured. And Fodor and Pylyshyn claim,

You don't, for example, get minds that are prepared to infer John went to the Store from John and Mary and Susan and Sally went to the store and from John and Mary went to the store but not from John and Mary and Susan went to the store. Given [linguistically structured representations], it is a truism that you don't get such minds. (1988, 48)

In short, human thought is inferentially coherent. Any example of inferential coherence is best explained by appeal to linguistically structured representations. Hence, inferential coherence in human thought is best explained by appeal to linguistically structured representations.

3. Problems and Objections

There are important problems for, and objections to, LOTH. The first is the problem of individuating the symbols of the language of thought, which if unsolvable would prove fatal for LOTH, at least so far as LOTH is to be a component of a fully naturalized theory of mind, or so far as it is to provide a framework within which psychological generalizations ranging across individuals may be made. The second is the problem of explaining context-dependent properties of thought, which should not exist if thinking is a computational process.

The third is the objection that contemporary cognitive science shows that some thinking takes place in mental images, which do not have a linguistic structure, so LOTH cannot be the whole story about rational thought. The fourth is the objection that systematicity, productivity, and inferential coherence may be accounted for in representational systems that do not employ linguistic formats (such as maps), so the arguments from those features do not prove LOTH. The fifth is the argument that connectionist networks, computational systems that do not employ linguistic representation, provide a more biologically realistic model of the human brain than do classical digital computers. The last part briefly raises the question whether the mind is best viewed as an analog or digital machine.

a. Individuating Symbols

An important and difficult problem concerning LOTH is the individuation of primitive symbols within the language of thought, the atomic mental representations. There are three possibilities for doing so: in terms of the meaning of a symbol, in terms of the syntax of a symbol (where syntactic kinds are conceived of as brain-state kinds), and in terms of the computational role of the symbol (for example, the causal relations the symbol bears to other symbols and to behavior).

Some authors (Aydde 1999 and Schneider 2009a) argue that this problem is perhaps fatal for LOTH. Schneider (2009a) argues that none of the above proposals (so far) are consistent with the roles that symbols are supposed to play within LOTH. In particular, an appeal to meaning in order to individuate symbols would not reduce intentionality to purely physical terms, and would thus stand opposed to a fully naturalized philosophy of mind. An appeal to syntax conceived of as brain states would amount to a type-identity theory for mental representation, and would thus be prone to difficulties faced by a general type-identity theory of mental states. An appeal to computational role would render impossible an explanation of how concepts can be shared by individuals, since no two individuals will employ symbols that have identical computational roles. A failure to explain how concepts may be shared, moreover, would render impossible the stating of true psychological generalizations ranging across individuals. See Schneider 2009b for a proposed solution to this problem.

b. Context-dependent Properties of Thought

Interestingly enough, Fodor himself has argued that LOTH (and CTM more generally), should be viewed as a thesis about a small portion of cognition. In his view, even were the theory to be completed, it would not offer an entire picture of the nature of thought (see Fodor 2000).

His primary argument for this conclusion is that computation is sensitive only to the syntax of the representations involved, so if thinking is computation it should be sensitive only to the syntax of mental representations, but quite often this appears not to be so. More specifically, the syntax of a representation is context-independent, but thoughts often have properties that are context-dependent.

For example, the thought it's raining might prompt the thought that's good, the garden needs it in the context of a dry spell, whereas it might prompt the thought maybe we'd better turn around in the context of a hike in the mountains. According to LOTH, however, the syntax of the thought it's raining is the same in both contexts, and according to CSMP, any computations involving that thought are sensitive only to its syntax. So there would seem to be no explanation why that thought would prompt different thoughts in different contexts, since the computations are not sensitive to those contexts. More generally, the role a given thought will play in one's thinking is a function of the entire body of propositions one believes. In Fodor's terminology, the complexity of a thought is not context-independent.

However, CTM would seem to require it to be. Thus according to Fodor, there is much cognition that cannot be understood on a computational model. See Ludwig and Schneider 2008 for an argument that this is not in fact a problem for LOTH.

c. Mental Images

http://www.iep.utm.edu/lot-hypo/print/
Throughout the 1970s, investigators designed a series of experiments concerned with mental imagery. The general conclusion many drew was that mental imagery presents a kind of mental representation that is not linguistically structured. More specifically, it was believed that the parts of mental images correspond to the spatial features of their content, whereas the parts of linguistic representations correspond to logical features of their content (see Kosslyn 1980).

In one well-known experiment, Kosslyn et al. (1978) asked subjects to memorize a map with various named locations on it. They then asked the subjects to imagine this map in their mind and to focus on a particular location. They asked the subjects (i) to say whether another given named location was on the map, and if so, (ii) to follow an imagined black dot as it traveled the shortest distance from the location on which they were focused to the named location (51). The result was that as the distance between the original location and the named location increased, so did the time it took subjects to respond. Kosslyn et al. concluded that "portions of images depict corresponding portions of the represented object(s) and that the spatial relations between portions of the imaged object(s) are preserved by the spatial relations between the corresponding portions of the image" (1978, 59-60).

It is important to note here that while the experiments involved invoke mental images as those images a subject can examine introspectively, the debate is best understood as being about non-introspectible mental representations. Since LOTH is a hypothesis about non-introspectible cognitive processing, any purported challenges to the hypothesis would likewise need to be about such processing. Thus if the above conclusion is correct, then it at least limits the scope of LOTH. Ned Block (1985) explains,

> The relevance of the pictorial/descriptional controversy to the viability of the computer metaphor in cognitive science should be becoming visible. The computer metaphor goes naturally with descriptional representations, but it is not at all clear how it can work when the representations are nondescriptional. (335)

However, some authors have denied that the data on mental imagery present a viable challenge to LOTH. Pylyshyn (1981) for instance, argues that the data are better explained by appeal to a kind of "tact knowledge" possessed by subjects or to architectural features of the cognitive system, but not to representations with non-linguistic structuring. Tye (1991) argues that on a proper understanding of the thesis that mental images have spatial properties, it does not (straightforwardly) undermine the claim that mental representation has a linguistic structure. Rather, he argues, it should be understood as the claim that mental images employ both non-linguistic and linguistic elements. See Block 1981 for a useful collection of essays on the imagery debate.

### d. Mental Maps

Another objection to LOTH comes from philosophers who have argued that there are non-linguistic forms of representation that are productive, systematic, and inferentially coherent. For example, David Braddon-Mitchell and Frank Jackson (1996) argue that maps are an important example. The point out that productivity, systematicity and inferential coherence show that thought must be structured, where a system of representation is structured just in case the similarities that hold between the representational states of the system reflect similarities that hold between the states that the system serves to represent, such that for new representational states, one can discover which states they serve to represent. They write,

> What is unbelievable is that similarities between the various [representational states] R should in no way correspond to similarities among the [represented states] S; it must be the case that enough information about a finite set of [R] giving which [S] each represents enables in principle the working out, for some new [R], which [S] it would represent. What it means to say that the way the R's serve to represent the S's is structured is that at some level of abstraction the similarities and differences between the R's correspond to similarities and differences among the S's, and it is this fact that underlies our ability to grasp for some new R which S it represents. (1996, 168-9)

They argue that maps are structured in this just sense, and can therefore account for productivity and systematicity (and presumably inferential coherence as well, but they do not argue for it). They point out that different parts of a map serve to represent different things (red dots for cities, blue lines for rivers, blue circles for lakes). Given these elements, there is no limit on the arrangement of ways in which a map may be constructed. Braddon-Mitchell and Jackson explain,

> the conventions of cartography do not set an upper limit on the number of different possible distributions of cities, areas of high pressure and the like that a map framed within those conventions can represent. A map-maker can represent quite a new situation as easily as a word- or sentence-maker can. (1996, 172-3)

They also argue that maps are systematic. They write,

> a map that represents Boston as being north of New York has the resources to represent New York as north of Boston, and a map that represented New York as north of Boston would be a kind of rearrangement of the map that represents Boston as north of New York. (1996, 172)

However, there are important differences between maps and linguistic representations. First, although maps have parts, they do not have atomic parts. As Braddon-Mitchell and Jackson put the point,

> There are many jigsaw puzzles you might make out of the map, but no single one would have a claim to have pieces that were all and only the most basic units. The reason is that there is no natural minimum unit of truth-assessable representation in the case of maps. (1996, 171)

Second, maps are "informationally rich" in the sense that they never express just a single proposition. Any map that expresses the proposition Boston is north of New York also expresses the proposition New York is south of Boston. One way to think about this difference is in terms of the smallest number of beliefs it is possible to have. For example, David Lewis (1994) questions whether, if thinking employs maps, the word 'belief' can be properly pluralized. He writes,

> No snippet of a map is big enough that, determinately, something is true according to it, and also small enough that, nothing is true according to any smaller part of it. If mental representation is map-like... then 'beliefs' is a bogue plural. You have beliefs the way you have the blues, or the mumps, or the shivers. But if mental representation is language-like, one belief is one sentence written in the belief-box, so 'beliefs' is a genuine plural. (311)
Third, the structuring that maps possess is of a different sort than the structuring possessed by linguistic representations. Specifically, the features of content that parts of maps correspond to are spatial features, whereas linguistic representations disregard spatial structure but correspond to logical features of content.

Hence, if the suggestion is that all thinking takes place in mental maps, then it presents a complete alternative to LOTH. This may be difficult to show, however, particularly for thought involving abstract concepts that are not easily expressible in map-form, though Braddon-Mitchell and Jackson do briefly offer one such argument (1996, 172). Camp (2007) argues that much, but not all, human thought may occur in maps, but that an organism of sufficiently limited cognitive capacity could think entirely in maps.

e. Connectionist Networks

The most widely discussed objection to LOTH is the objection that connectionist networks provide better models of cognition than computers processing linguistically structured representations (see Bechtel and Abramson 1990, Churchland 1984, and Elman et al. 1996 for useful introductions). Such networks possess some number of interconnected nodes, typically arranged as layers of input, output, and hidden nodes. Each node possesses a level of activation, and each connection is weighted. The level of activation of all the nodes to which a given node is connected, together with the weightings of those connections, determine the level of activation of the given node. A particular set of activations at the input nodes will result in a particular set of activations at the output nodes.

The activation of a given set of nodes (typically input layers and output layers) can be interpreted as having semantic content, but the activation level of a particular node can not. Moreover, the interpretation of the activations of a set of nodes does not result from the collection of activations of the particular nodes involved in anything like the way the semantic content of a linguistically structured compound representation results from the content of its component parts (that is, they do not combine via concatenation). In short, connectionist networks possess neither combinatorial syntax nor compositional semantics; the representations involved are not linguistically structured.

There are, however, many ways in which networks resemble the brain and its functioning more closely than do digital computers (the canonical model of a linguistic representation processor). The most obvious is that the brain is a massive network of neurons, as connectionist machines are networks of nodes, and does not possess a central processing unit, as do digital computers. Moreover, processing in both the brain and connectionist networks is distributed and parallel, while it is serial in digital computers and concentrated in the central processing unit. Activation levels in both nodes and neurons are defined by continuous numerical values, while representations in digital machines are discrete elements, and processing takes place in discrete steps. It is for these and similar reasons that connectionists have taken networks to offer more “biologically realistic” models of the mind than the digital computer. Smolensky (1988) is careful to note however, that connectionist networks also differ from the brain in many important respects (for example, nodes in a network are uniformly dense, while neurons are more highly connected to neighboring neurons) and thus that the notion that they are “biologically realistic” can be misleading and should be treated with caution.

Much of the debate concerning connectionist networks is about whether or not they provide a real alternative to LOTH. In particular, it is agreed that networks can implement systems that process linguistically structured representations. Such networks may provide useful models of cognition at a level of analysis below the level at which LOTH operates—that is, they may provide an analysis of how higher cognition is implemented in the brain. The question then, is whether they can offer an alternative to LOTH itself, which purports to explain how such (supposed) higher features of cognition such as productivity, systematicity, and inferential coherence, are possible. If they can explain these features without implementing a system that processes linguistically structured representations, then they do indeed offer an alternative to LOTH.

Smolensky (1987) argues that representations in (some) networks do have adequate constituent structure to account for such features as systematicity and inferential coherence. For instance, he suggests that a representation of the concept cup with coffee would include various “microfeatures” (hot liquid, burnt odor, and so forth) that are not included in a representation of the concept cup without coffee. These microfeatures, then, not only comprise a constituent of the representation, but would also comprise a representation of the concept coffee. However, Smolensky admits that these sorts of constituents may not be exact copies of each other in different contexts, but rather will bear a “family resemblance” to one another, such that the features they share are enough to produce “common processing implications.” Fodor and McLaughlin (1990) argue in response that only constituency as it occurs in linguistically structured representations (in which constituents of a representation are tokened whenever the representation in tokened, and in which those constituents are identical across varying contexts) can account for systematicity and inferential coherence, and so Smolensky’s account of constituency in networks cannot explain those features. See Horgan and Tienson 1991 for a useful collection of papers on connectionism and its relation to LOTH.

f. Analog and Digital Representation

One commonality that holds among the last three objections discussed is that they can all reasonably be described as claiming that at least some mental representation is analog, while LOTH describes mental representation as digital. The distinction is usually understood in terms of continuity and discreteness. Digital representations are discrete (as words and sentences). Analog representations are continuous, or possess continuously variable properties such as distances between parts of an image or map, or activation values of the nodes in a network.

However, the distinction between analog and digital representation has been understood in a number of ways. David Lewis (1971) says that “analog representation of numbers is representation of numbers by physical magnitudes that are either primitive or almost primitive,” (325) and that “digital representation of numbers [is] representation of numbers by differentiated multi-digital magnitudes” (327). Fred Dretske (1981) says that “a signal... carries the information that s is F in digital form if and only if the signal carries no additional information about s, no information that is not already nested in s’s being F. If the signal does carry additional information about s, information that is not nested in s’s being F, then... the signal carries this information in analog form (157). And James Blachowicz (1997) says that “the function of analog representation is to map or model what it represents (83). See also Von Neumann 1958, Goodman 1968, Trenholme 1994, Haugeiland 1998, and Katz 2008.

The analog/digital distinction may be drawn in reference to different kinds of things: computers, representations, processes, machines, and so forth. Haugeland (1998) argues that, although all digital representations share some important features, there may be no set of
features uniquely characterizing analog representation. If that is the case, then the idea that images, maps, and networks are analog should not be taken to indicate that they share some important set of features other than being non-digital. Moreover, because it remains a possibility that thought is best modeled by a connectionist network implementing a system that processes linguistically structured representations, and because it remains a possibility that some thinking takes place in images, some in maps, some in linguistically structured representations, and some in yet other forms of representation, it would be misleading to suggest that the question whether the mind is best modeled by an analog or digital machine has a singular answer.

4. References and Further Reading


Author Information

Matthew Katz
Email: katzma@cmich.edu
Central Michigan University