STANFORD ENCYCLOPEDIA OF PHILOSOPHY

The Language of Thought Hypothesis

First published Thu May 28, 1998; substantive revision Fri Sep 17, 2010

The Language of Thought Hypothesis (LOTH) postulates that thought and thinking take place in a mental language. This language consists of a system of representations that is physically realized in the brain of thinkers and has a combinatorial syntax (and semantics) such that operations on representations are causally sensitive only to the syntactic properties of representations. According to LOTH, thought is, roughly, the tokening of a representation that has a syntactic (constituent) structure with an appropriate semantics. Thinking thus consists in syntactic operations defined over such representations. Most of the arguments for LOTH derive their strength from their ability to explain certain empirical phenomena like productivity and systematicity of thought and thinking.

1. What is the Language of Thought Hypothesis?

LOTH is an empirical thesis about the nature of thought and thinking. According to LOTH, thought and thinking are done in a mental language, i.e., in a symbolic system physically realized in the brain of the relevant organisms. In formulating LOTH, philosophers have in mind primarily the variety of thoughts known as 'propositional attitudes'. Propositional attitudes are the thoughts described by such sentence forms as '*S* believes that *P*', '*S* hopes that *P*', '*S* desires that *P*', etc., where '*S*' refers to the subject of the attitude, '*P*' is any sentence, and 'that *P*' refers to the proposition that is the object of the attitude. If we let '*A*' stand for such attitude verbs as 'believe', 'desire', 'hope', 'intend', 'think', etc., then the propositional attitude statements all have the form: *S A*s that *P*.

LOTH can now be formulated more exactly as a hypothesis about the nature of propositional attitudes and the way we entertain them. It can be characterized as the conjunction of the following three theses (A), (B) and (C):

A. Representational Theory of Mind (RTM) (cf. Field 1978:37, Fodor 1987:17):

    1. Representational Theory of Thought: For each propositional attitude *A*, there is a unique and distinct (i.e. dedicated)[1] psychological relation *R*, and for all propositions *P* and subjects *S*, *S A*s that *P* if and only if there is a mental representation *#P#* such that

        a. *S* bears *R* to *#P#*, and
        b. *#P#* means that *P*.

    2. Representational Theory of Thinking: Mental processes, thinking in particular, consists of causal sequences of tokenings of mental representations.

B. Mental representations, which, as per (A1), constitute the direct "objects" of propositional attitudes, belong to a representational or symbolic *system* which is such that (cf. Fodor and Pylyshyn 1988:12–3)

    1. representations of the system have a combinatorial syntax and semantics: structurally complex (molecular) representations are systematically built up out of structurally simple (atomic) constituents, and the semantic content of a molecular representation is a function of the semantic content of its atomic

constituents together with its syntactic/formal structure, *and*

2. the operations on representations (constituting, as per (A2), the domain of mental processes, thinking) are causally sensitive to the syntactic/formal structure of representations defined by this combinatorial syntax.

C. Functionalist Materialism. Mental representations so characterized are, at some suitable level, functionally characterizable entities that are (possibly, multiply) realized by the physical properties of the subject having propositional attitudes (if the subject is an organism, then the realizing properties are presumably the neurophysiological properties of the brain).

The relation *R* in (A1), when RTM is combined with (B), is meant to be understood as a *computational/functional* relation. The idea is that each attitude is identified with a characteristic computational/functional role played by the mental sentence that is the direct "object" of that kind of attitude. (Scare quotes are necessary because it is more appropriate to reserve 'object' for a *proposition* as we have done above, but as long as we keep this in mind, it is harmless to use it in this way for LOT sentences.) For instance, what makes a certain mental sentence an (occurrent) belief might be that it is characteristically the output of perceptual systems and input to an inferential system that interacts decision-theoretically with desires to produce further sentences or action commands. Or equivalently, we may think of belief sentences as those that are accessible only to certain sorts of computational operations appropriate for beliefs, but not to others. Similarly, desire-sentences (and sentences for other attitudes) may be characterized by a different set of operations that define a characteristic computational role for them. In the literature it is customary to use the metaphor of a "belief-box" (cf. Schiffer 1981) as a blanket term to cover whatever specific computational role belief sentences turn out to have in the mental economy of their possessors. (Similarly for "desire-box", etc.)

The Language of Thought Hypothesis is so-called because of (B): token mental representations are like sentences in a language in that they have a syntactically and semantically regimented constituent structure. Put differently, the mental representations that are the direct "objects" of attitudes are structurally complex symbols whose complexity lends itself to a syntactic and semantic analysis. This is also why the LOT is sometimes called *Mentalese*.

It is (B2) that makes LOTH a species of the so-called Computational Theory of Mind (CTM). This is why LOTH is sometimes called the Computational/Representational Theory of Mind or Thought (CRTM/CRTT) (cf. Rey 1991, 1997). Indeed, LOTH seems to be the most natural product when RTM is combined with a view that treats mental processes or thinking as computational when computation is understood traditionally or *classically* (this is a recent term emphasizing the contrast with connectionist processing, which we will discuss later).

According to LOTH, when someone believes that *P*, there is a sense in which the immediate "object" of one's belief can be said to be a complex symbol, a sentence in one's LOT physically realized in the neurophysiology of one's brain, that has both syntactic structure and a semantic content, namely the proposition that *P*. So, contrary to the orthodox view that takes the belief relation as a dyadic relation between an agent and a proposition, LOTH takes it to be a triadic relation among an agent, a Mentalese sentence, and a proposition. The Mentalese sentence can then be said to have the proposition as its semantic/intentional content. Within the framework of LOTH, it is only in this sense can it be said that what is believed is a proposition, and thus the proper *object* of the attitude.

This triadic view seems to have several advantages over the orthodox dyadic view. It is a puzzle in the dyadic view how intentional organisms can stand in direct relation to abstract objects like propositions in such a way as to influence their causal powers. According to *folk psychology* (ordinary commonsense psychology that we rely on daily in our dealings with others), it is because those states have the propositional content they do that they have the causal powers they do. LOTH makes this relatively non-mysterious by introducing a physical intermediary that is capable of having the relevant causal powers in virtue of its syntactic structure that encodes its semantic content. Another advantage of this is that the thought processes can be causally guided by the syntactic forms of the sentences in a way that respect their semantic contents. This is the virtue of (B) to which we'll come back below. Mainly because of these features, LOTH is said to be poised to scientifically vindicate folk psychology if it turns out to be true.

2. Status of LOTH

LOTH has primarily been advanced as an *empirical* thesis (although some have argued for the truth of LOTH on a priori or conceptual grounds following the natural conceptual contours of folk psychology—see Davies 1989, 1991; Lycan 1993; Rey 1995; Jacob 1997; Markic 2001 argues against Jacob. Harman 1973 develops and defends LOTH on both empirical and conceptual grounds). It is not meant to be taken as an analysis of what the folk *mean* (or, for that matter, what the scientists ought to mean) when they talk about various propositional attitudes and their role in thinking. In this regard, LOT theorists typically view themselves as engaged in some sort of a proto-science, or at least in some empirical research program continuous with scientific psychology. Indeed, as we will see in more detail below, when Jerry Fodor first explicitly articulated and elaborated LOTH in some considerable detail in his (1975), he basically defended it on the ground that it was assumed by our best scientific theories or models in cognitive psychology and psycholinguistics. This empirical status generally accorded to LOTH should be kept firmly in mind when assessing its plausibility and especially its prospects in the light of new evidence and developments in scientific psychology. Nevertheless, it would be more appropriate to see LOTH more as a foundational thesis rather than as an ongoing research project guided by a set of concrete empirical methods, specific theses and principles. In this regard, LOTH stands to specific scientific theories of the (various aspects of the) mind somewhat like the "Atomic Hypothesis" stands to a whole bunch specific scientific theories about the particulate nature of the world (some of which may be—and certainly historically, have been—incompatible with each other).

When viewed this way, scientific theories advanced within the LOTH framework are not, strictly speaking, committed to preserving the folk taxonomy of the mental states in any very exact way. Notions like belief, desire, hope, fear, etc. are folk notions and, as such, it may not be utterly plausible to expect (eliminativist arguments aside) that a scientific psychology will preserve the exact contours of these concepts. On the contrary, there is every reason to believe that scientific counterparts of these notions will carve the mental space somewhat differently. For instance, it has been noted that the folk notion of belief harbors many distinctions. For example, it has both a dispositional and an occurrent sense. In the occurrent sense, it seems to mean something like consciously entertaining and accepting a thought (proposition) as true. There is quite a bit of literature and controversy on the dispositional sense.[2] Beliefs are also capable of being explicitly stored in long term memory as opposed to being merely dispositional or tacit. Compare, for instance: I believe that there was a big surprise party for my 24th birthday vs. I have always believed that lions don't eat their food with forks and knives, or that 13652/4=3413, even though until now these latter two thoughts had never occurred to me. There is furthermore the issue of degree of belief: while I may believe that George will come to dinner with his new girlfriend even though I wouldn't bet on it, you, thinking that you know him better than I do, may nevertheless go to the wall for it. It is unlikely that there will be one single construct of scientific psychology that will exactly correspond to the folk notion of belief in all these ways.

For LOTH to vindicate folk psychology it is sufficient that a scientific psychology with a LOT architecture come up with scientifically grounded psychological states that are recognizably like the propositional attitudes of folk psychology, and that play more or less similar roles in psychological explanations.[3]

3. Scope of LOTH

LOTH is an hypothesis about the nature of thought and thinking with propositional content. As such, it may or may not be applicable to other aspects of mental life. Officially, it is silent about the nature of some mental phenomena such as experience, qualia,[4] sensory processes, mental images, visual and auditory imagination, sensory memory, perceptual pattern-recognition capacities, dreaming, hallucinating, etc. To be sure, many LOT theorists hold views about these aspects of mental life that sometimes make it seem that they are also to be explained by something similar to LOTH.[5]

For instance, Fodor (1983) seems to think that many modular input systems have their own LOT to the extent to which they can be explained in representational and computational terms. Indeed, many contemporary psychological models treat perceptual input systems in just these terms.[6] There is indeed some evidence that this kind of treatment might be appropriate for many perceptual processes. But it is to be kept in mind that a system may employ representations and be computational without necessarily satisfying any or both of the clauses in (B) above in any full-fledged way. Just think of finite automata theory where there are plenty of examples of a computational process defined over states or symbols which lack full-blown syntactic and/or semantic structural complexity. (For a useful discussion of varieties of computational processes and their classification, see Piccinini 2008.) Whether sensory or perceptual processes are to be treated within the framework of full-blown LOTH is again an open empirical question. It might be that the answer to this question is affirmative. If so, there may be more than one LOT realized in different subsystems or mechanisms in the mind/brain. So LOTH is not committed to there being a single representational system realized in the brain, nor is it committed to the claim that all mental representations are complex or language-like, nor would it be falsified if it turns out that most aspects of mental life other than the ones involving propositional attitudes don't require a LOT.

Similarly, there is strong evidence that the mind also exploits an image-like representational medium for certain kinds of mental tasks.[7] LOTH is non-committal about the existence of an image-like representational system for many mental tasks other than the ones involving propositional attitudes. But it *is* committed to the claim that propositional thought and thinking cannot be successfully accounted for in its entirety in purely imagistic terms. It claims that a combinatorial sentential syntax is necessary for propositional attitudes and a purely imagistic medium is not adequate for capturing that.[8]

There are in fact some interesting and difficult issues surrounding these claims. The adequacy of an imagistic system seems to turn on the nature of syntax at the *sentential* level. For instance, Fodor, in Chapter 4 of his (1975) book, allows that many lexical items in one's LOT may be image-like; he introduces the notion of a *mental image/picture under description* to avoid some obvious inadequacies of pictures (e.g., what makes a picture a picture of an overweight woman rather than a pregnant one, or vice versa, etc.). This is an attempt to combine discursive and imagistic representational elements at the *lexical* level. There may even be a well defined sense in which pictures can be combined to produce structurally complex pictures (as in British Empiricism: image-like simple ideas are combined to produce complex ideas, e.g., the idea of a unicorn—see also Prinz 2002). But what is absolutely essential for LOTH, and what Fodor insists on, is the claim that there is no adequate way in which a purely image-like system can capture what is involved in making *judgments*, i.e., in judging *propositions* to be true. This seems to require a discursive syntactic approach at the sentential level. The general problem here is the inadequacy of pictures or image-like representations to express propositions. I can judge that the blue box is on top of the red one without judging that the red box is under the blue one. I can judge that Mary kisses John without judging that John kisses Mary, and so on for indefinitely many such cases. It is hard to see how images or pictures can do that without using any syntactic structure or discursive elements, to say nothing of judging, e.g., conditionals, disjunctive or negative propositions, quantifications, negative existentials, etc.[9]

Moreover, there are difficulties with imagistic representations arising from demands on *processing* representations. As we will see below, (B2) turns out to provide the foundations for one of the most important arguments for LOTH: it makes it possible to mechanize thinking understood as a semantically coherent thought process, which, as per (A2), consists of a causal sequence of tokenings of mental representations. It is not clear, however, how an equivalent of (B2) could be provided for images or pictures in order to accommodate operations defined over them, even if something like an equivalent of (B1) could be given. On the other hand, there are truly promising attempts to *integrate* discursive symbolic theorem-proving with reasoning with image-like symbols. They achieve impressive efficiency in theorem-proving or in any deductive process defined over the expressions of such an integrated system. Such attempts, if they prove to be generalizable to psychological theorizing, are by no means threats to LOTH; on the contrary, such systems have every feature to make them a species of a LOT system: they satisfy (B).[10]

4. Nativism and LOTH

In the book (1975) in which Fodor introduced the LOTH, he also argued that all concepts are innate. As a result, the connection between LOTH and an implausibly strong version of conceptual nativism looked very much internal. This historical coincidence has led some people to think that LOTH is essentially committed to a very strong form of nativism, so strong in fact that it seems to make a *reductio* of itself (see, for instance, P.S. Churchland 1986, H. Putnam 1988, A. Clark 1994). The gist of his argument was that since learning concepts is a form of hypothesis formation and confirmation, it requires a system of mental representations in which formation and confirmation of hypotheses are to be carried out, but then there is a non-trivial sense in which one already has (albeit potentially) the resources to express the extension of the concepts to be learned.

In his *LOT 2* (2008), Fodor continues to claim that concepts cannot be learned and that the very idea of concept learning is “confused”:

> Now, according to HF [the Hypothesis Formation and Confirmation model], the process by which one learns C must include the inductive evaluation of some such hypothesis as ‘The C things are the ones that are green or triangular’. But the inductive evaluation of that hypothesis itself requires (*inter alia*) bringing the property *green or triangular* before the mind as such. ... Quite generally, you can't represent anything as *such and such* unless you already have the concept *such and such*. All that being so, it follows, on pain of circularity, that ‘concept learning’ as HF understands it *can't* be a way of acquiring concept C. ... Conclusion: *If concept learning is as HF understands it, there can be no such thing*. This conclusion is entirely general; it doesn't matter whether the target concept is primitive (like GREEN) or complex (like GREEN OR TRIANGULAR). (*LOT 2*, 2008:139)

Note that this argument and the predecessors Fodor articulated in his previous writings and especially in his (1975) are entirely general, applicable to any hypothesis that identifies concepts with mental representations whether or not these representations belong to a LOT.

The crux of the issue seems to be that learning concepts is a rational process. There seem to be non-arbitrary semantic and epistemic liaisons between the target concept to be acquired and its “evidence” base. This evidence base needs to be represented and rationally tied to the target concept. This target concept needs also to be expressed in terms of representations one already possesses. Fodor thinks that any model of concept *learning* understood in this sense will have to be a form of hypothesis formation and confirmation. But not every form of concept *acquisition* is learning. There are non-rational ways of acquiring concepts whose explanation need not be at the cognitive level (e.g., brute triggering mechanisms that can be activated in sorts of ways that can presumably be explained at the sub-cognitive or neurophysiological levels). If concepts cannot be learned, then they are either innate or non-rationally acquired. Whereas early Fodor used to think that concepts must therefore be innate (maybe he thought that non-learning concept acquisition forms are limited to sensory or certain classes of perceptual concepts), he now thinks that they may be acquired but the explanation of this is not the business of cognitive psychology.

Whatever one may think of the merits of Fodor's arguments for concept nativism or of his recent anti-learning stance, it should be emphasized that LOTH per se has very little to do with it. LOTH is not committed to such a strong version of nativism, especially about concepts. It also need not be committed to any anti-learning stance about concepts. It is certainly plausible to assume that LOTH will turn out to have some empirically (as well as theoretically/a priori) motivated nativist commitments about the structural organization and dynamic management of the entire representational system. But this much is to be expected especially in the light of recent empirical findings and trends. This, however, does not constitutes a *reductio*. It is an open empirical question how much nativism is true about concepts, and LOTH should be so taken as to be capable of accommodating whatever turns out to be true in this matter. LOTH, therefore, when properly conceived, is independent of any specific proposal about *conceptual* nativism.[11]

5. Naturalism and LOTH

One of the most attractive features of LOTH is that it is a central component of an ongoing research program in philosophy of psychology to naturalize the mind, that is, to give a theoretical framework in which the mind could naturally be seen as part of the physical world without postulating irreducibly psychic entities, events, processes or properties. Fodor, historically the most important defender of LOTH, once identified the major mysteries in philosophy of mind thus:

> How could anything material have conscious states? How could anything material have semantical properties? How could anything material be rational? (where this means something like: how could the state transitions of a physical system preserve semantical properties?). (1991: 285, Reply to Devitt)

LOTH is a full-blown attempt to give a naturalist answer to the third question, an attempt to solve at least part of the problem underlying the second one, and is almost completely silent about the first.[12]

According to RTM, propositional attitudes are relations to meaningful mental representations whose causally sequenced tokenings constitute the process of thinking. This much can, in principle, be granted by an intentional realist who might nevertheless reject LOTH. Indeed, there are plenty of theorists who accept RTM in some suitable form (and also happily accept (C) in many cases) but reject LOTH either by explicitly rejecting (B) or simply by remaining neutral about it. Among some of the prominent philosophers who choose the former option are Searle (1984, 1990, 1992), Stalnaker (1984), Lewis (1972), Barwise and Perry (1983).[13] Some who want to remain neutral include Loar (1982a, 1982b), Dretske (1981), Armstrong (1980), and many contemporary functionalists including some connectionists.[14]

But RTM per se doesn't so much propose a naturalistic solution to intentionality and mechanization of thinking as simply assert a framework to emphasize intentional realism and, perhaps, with (C), a declaration of a commitment to naturalism or physicalism at best. How, then, is the addition of (B) supposed to help? Let us first try to see in a bit more detail what the problem is supposed to be in the first place to which (B) is proposed as a solution. Let us start by reflecting on thinking and see what it is about thinking that makes it a mystery in Fodor's list. This will give rise to one of the most powerful (albeit still nondemonstrative) arguments for LOTH.

5.1 The Problem of Thinking

RTM's second clause (A2), in effect, says that thinking is *at least* the tokenings of states that are (a) intentional (i.e. have representational/propositional content) and (b) causally connected. But, surely, thinking is more. There could be a causally connected series of intentional states that makes no sense at all. Thinking, therefore, is causally proceeding from states to states that makes semantic sense: the transitions among states must preserve some of their semantic properties to count as thinking. In the ideal case, this property would be the truth value of the states. But in most cases, any interesting intentional or epistemic property would do (e.g., warrantedness, degree of confirmation, semantic coherence given a certain practical context like satisfaction of goals in a specific context, etc.). In general, it is hard to spell out what this requirement of “making sense” comes to. The intuitive idea, however, should be clear. Thinking is not proceeding from thoughts to thoughts in arbitrary fashion: thoughts that are causally connected are in some fashion semantically (rationally, epistemically) connected too. If this were not so, there would be little point in thinking—thinking couldn't serve any useful purpose. Call this general phenomenon, then, the *semantic coherence* of causally connected thought processes. LOTH is offered as a solution to this puzzle: how is thinking, conceived this way, physically possible? This is the problem of thinking, thus the problem of mechanization of rationality in Fodor's version. How does LOTH propose to solve this problem and bring us one big step closer to the naturalization of the mind?

5.2 Syntactic Engine Driving a Semantic Engine: Computation

The two most important achievements of 20th century that are at the foundations of LOTH as well as most of modern Artificial Intelligence (AI) research and most of the so-called information processing approaches to cognition are (i) the developments in modern symbolic (formal) logic, and (ii) Alan Turing's idea of a Turing Machine and Turing computability. It is putting these two ideas together that gives LOTH its enormous explanatory power within a naturalistic framework. Modern logic showed that most of deductive reasoning can be formalized, i.e. most semantic relations among symbols can be entirely captured by the symbols' formal/syntactic properties and the relations among them. And Turing showed, roughly, that if a process has a formally specifiable character then it can be mechanized. So we can appreciate the implications of (i) and (ii) for the philosophy of psychology in this way: if thinking consists in processing representations physically realized in the brain (in the way the internal data structures are realized in a computer) and these representations form a formal system, i.e., a language with its proper combinatorial syntax (and semantics) and a set of derivations rules formally defined over the syntactic features of those representations (allowing for specific but powerful programs to be written in terms of them), then the problem of thinking, as described above, can in principle be solved in completely naturalistic terms, thus the mystery surrounding how a physical device can ever have semantically coherent state transitions (processes) can be removed. Thus, given the commitment to naturalism, the hypothesis that the brain is a kind of computer trafficking in representations in virtue of their syntactic properties is the basic idea of LOTH (and the AI vision of cognition).

Computers are environments in which symbols are manipulated in virtue of their formal features, but what is thus preserved are their semantic properties, hence the semantic coherence of symbolic processes. Slightly paraphrasing Haugeland (cf. 1985: 106), who puts the same point nicely in the form of a motto:

> *The Formalist Motto*:
> If you take care of the syntax of a representational system, its semantics will take care of itself.

This is in virtue of the mimicry or mirroring relation between the semantic and formal properties of symbols. As Dennett once put it in describing LOTH, we can view the thinking brain as a syntactically driven engine preserving semantic properties of its processes, i.e. driving a semantic engine. What is so nice about this picture is that if LOTH is true we have a naturalistically adequate causal treatment of *thinking* that respects the semantic properties of the *thoughts* involved: it is in virtue of the physically coded syntactic/formal features that thoughts cause each other while the coherence of their semantic properties is preserved precisely in virtue of this.

Whether or not LOTH actually turns out to be empirically true in the details or in its entire vision of rational thinking, this picture of a syntactic engine driving a semantic one can at least be taken to be an important *philosophical* demonstration of how Descartes' challenge can be met (cf. Rey 1997: chp.8). Descartes claimed that rationality in the sense of having the power "to act in all the contingencies of life in the way in which our reason makes us act" cannot possibly be possessed by a purely physical device: "The rational soul … could not be in any way extracted from the power of matter … but must … be expressly created" (1637/1970: 117–18). Descartes was completely puzzled by just this rational character and semantic coherence of thought processes so much so that he failed to even imagine a possible mechanistic explication of it. He thus was forced to appeal to Divine creation. But we can now see/imagine at least a possible mechanistic/naturalistic scenario.[15]

5.3 Intentionality and LOTH

But where do the semantic properties of the mental representations come from in the first place? How can they mean anything? This is Brentano's challenge to a naturalist. Brentano's bafflement was with the intentionality of the human mind, its apparently mysterious power to represent things, events, properties in the world. He thought that nothing physical can have this property: "The reference to something as an object is a distinguishing characteristic of all mental phenomena. No physical phenomenon exhibits anything similar" (Brentano 1874/1973: 97). This problem of intentionality is the second problem or mystery in Fodor's list quoted above. I said that LOTH officially offers only a partial solution to it and perhaps proposes a framework within which the remainder of the solution can be couched and elaborated in a naturalistically acceptable way.

Recall that RTM contains a clause (A1b) that says that the immediate "object" of a propositional attitude that *P* is a mental representation #*P*# that *means* that *P*. Again, (B1) attributes a compositional semantics to the syntactically complex symbols belonging to one's LOT that are, as per (C), realized by the physical properties of a thinking system. According to LOTH, the semantic content of propositional attitudes is inherited from the semantic content of the mental symbols. So Brentano's questions for a LOT theorist becomes: how do the symbols in one's LOT get their meanings in the first place? There are two levels or stages at which this question can be raised and answered:

> (1) At the level of *atomic* symbols (non-logical primitives): how do the atomic symbols represent what they do?
>
> (2) At the level of *molecular* symbols (phrasal complexes or sentences): how do molecular symbols represent what they do?

There have been at least two major lines LOT theorists have taken regarding these questions. The one that is least committal might perhaps be usefully described as the *official position* regarding LOTH's treatment of intentionality. Most LOT theorists seem to have taken this line. The official line doesn't propose any theory about the first stage, but simply assumes that the first question can be answered in a naturalistically acceptable way. In other words, officially LOTH simply assumes that the atomic symbols/expressions in one's LOT have whatever meanings they have.[16]

But, the official line continues, LOTH has a lot to say about the second stage, the stage where the semantic contents are computed or assigned to complex (molecular) symbols on the basis of their combinatorial syntax or grammar together with whatever meanings atomic symbols are assumed to have in the first stage. This procedure is familiar from a Tarski-style[17] definition of truth conditions of *sentences*. The truth-value of complex sentences in propositional logic are completely determined by the truth-values of the atomic sentences they contain together with the rules fixed by the truth-tables of the connectives occurring in the complex sentences. Example: ‘*P* and *Q*’ is true just in case both ‘*P*’ and ‘*Q*’ are true, but false otherwise. This process is similar but more complex in first-order languages, and even more so for natural languages—in fact, we don't have a completely working compositional semantics for the latter at the moment. So, *if* we have a semantic interpretation of atomic symbols (*if* we have symbols whose reference and extension are fixed at the first stage by whatever naturalistic mechanism turns out to govern it), *then* the combinatorial syntax will take over and effectively determine the semantic interpretation (truth-conditions) of the complex sentences they are constituents of. So officially LOTH would only contribute to a complete naturalization project if there is a naturalistic story at the atomic level.

Early Fodor (1975, 1978, 1978a, 1980), for instance, envisaged a science of psychology which, among other things, would reasonably set for itself the goal of discovering the combinatorial syntactic principles of LOT and the computational rules governing its operations, without worrying much about semantic matters, especially about how to fix the semantics of atomic symbols (he probably thought that this was not a job for LOTH). Similarly, Field (1978) is very explicit about the combinatorial rules for assigning truth-conditions to the sentences of the internal code. In fact, Field's major argument for LOTH is that, given a naturalistic causal theory of reference for atomic symbols, about which he is optimistic (Field 1972), it is the only naturalistic theory that has a chance of solving Brentano's puzzle. For the moment, this is not much more than a hope, but, according to the LOT theorist, it is a well-founded hope based on a number of theoretical and empirical assumptions and data. Furthermore, it is a framework defining a naturalistic research program in which there have been promising successes.[18]

As I said, this official and, in a way, least committal line has been the more standard way of conceiving LOTH's role in the project of naturalizing intentionality. But some have gone beyond it and explored the ways in which the resources of LOTH can be exploited even in answering the first question (1) about the semantics of atomic symbols.

Now, there is a weak version of an answer to (1) on the part of LOTH and a strong version. On the weak version, LOTH may be untendentiously viewed as inevitably providing *some* of the resources in giving the ultimate naturalistic theory in naturalizing the meaning of atomic symbols. The basic idea is that whatever the ultimate naturalistic theory turns out to be true about atomic expressions, computation as conceived by LOTH will be part of it. For instance, it may be that, as with nomic covariation theories of meaning (Fodor 1987, 1990a; Dretske 1981), the meaning of an atomic predicate may consist in its potential to get tokened in the presence of (or, in causal response to) something that instantiates the property the predicate is said to express. A natural way of explicating this potential may partly but ultimately rely on certain computational principles the symbol may be subjected to within a LOT framework, or principles that in some sense govern the “behavior” of the symbol. Insofar as computation is naturalistically understood in the way LOTH proposes, a complete answer to the first question about the semantics of atomic symbols may plausibly involve an explicatory appeal to computation within a system of symbols. This is the weak version because it doesn't see LOTH as proposing a complete solution to the first question (1) above, but only *helping* it.

A strong version would have it that LOTH provides a *complete* naturalistic solution to both questions: given the resources of LOTH we don't need to look any further to meet Brentano's challenge. The basic idea lies in so-called functional or conceptual role semantics, according to which a concept is the concept it is precisely in virtue of the particular causal/functional potential it has in interacting with other concepts. Each concept may be thought of as having a certain distinctive set of epistemic/semantic relations or liaisons to other concepts. We can conceive of this set as determining a certain “conceptual role” for each concept. We can then take these roles to determine the semantic identity of concepts: concepts are the concepts they are because they have the conceptual roles they have; that is to say, among other things, concepts represent whatever they do precisely in virtue of these roles. The idea then is to reduce each *conceptual* role to *causal/functional* role of atomic symbols (now conceived as primitive terms in LOTH), and then use the resources of LOTH to reduce it in turn to *computational* role. Since computation is naturalistically well-defined, the argument goes, and since causal interactions between thoughts and concepts can be understood completely in terms of computation, we can completely naturalize intentionality if we can successfully treat meanings as arising out of thoughts/concepts' internal interactions with each other. In other words, the strong version of LOTH would claim that atomic symbols in LOT have the content they do in virtue of their potential for causal interactions with other tokens, and cashing out this potential in mechanical/naturalistic terms is what, among other things, LOTH is for. LOTH then comes as a naturalistic rescuer for conceptual role semantics.

It is not clear whether any one holds this strong version of LOTH in this rather naive form. But certainly some people have elaborated the basic idea in quite subtle ways, for which Cummins (1989: chp.8) is perhaps the best example. (But also see Block 1986 and Field 1978.) But even in the best hands, the proposal turns out to be very problematic and full of difficulties nobody seems to know how to straighten out. In fact, some of the most ardent critics of taking LOTH as incorporating a functional role semantics turn out to be some of the most ardent defenders of LOTH understood in a weak, non-committal sense we have explored above—see Fodor (1987: chp.3), Fodor and Lepore (1991), Fodor's attack (1978b) on AI's way of doing procedural semantics is also relevant here. Haugeland (1981), Searle (1980, 1984), and Putnam (1988) quite explicitly take LOTH to involve a program for providing a complete semantic account of mental symbols, which they then attack accordingly.[19]

It is also possible, in fact, quite natural, to combine conceptual role semantics (internalist) with causal/informational psychosemantics (externalist). The result is sometimes known as two-factor theories. If this turns out to be the right way to naturalize intentionality, then, given what is said above about the potential resources of LOTH in contributing to both factors, it is easy to see why many theorists who worry about naturalizing intentionality are attracted to LOTH.

As indicated previously, LOTH is almost completely silent about consciousness and the problem of qualia, the third mystery in Fodor's list in the quote above. But the naturalist's hope is that this problem too will be solved, if not by LOTH, then by something else. On the other hand, it is important to emphasize that LOTH is neutral about the naturalizability of consciousness/qualia. If it turns out that qualia cannot be naturalized, this would by no means show that LOTH is false or defective in some way. In fact, there are people who *seem* to think that LOTH may well turn out to be true even though qualia can perhaps not be naturalized (e.g., Block 1980, Chalmers 1996, McGinn 1991).

Finally, it should be emphasized that LOTH has no particular commitment to every symbolic activity's being conscious. Conscious thoughts and thinking may be the tip of a computational iceberg. Nevertheless, there are ways in which LOTH can be helpful for an account of state consciousness that seeks to explain a thought's being conscious in terms of a higher order thought which is about the first order thought. So, to the extent to which thought and thinking are conscious, to that extent LOTH can perhaps be viewed as providing some of the necessary resources for a naturalistic account of state consciousness —for elaboration see Rosenthal (1997) and Lycan (1997).

6. Arguments for LOTH

We have already seen two major arguments, perhaps historically the most important ones, for LOTH: First, we have noted that if LOTH is true then all the essential features of the common sense conception of propositional attitudes will be explicated in a naturalistic framework which is likely to be co-opted by scientific cognitive psychology, thus vindicating folk psychology. Second, we have discussed that, if true, LOTH would solve one of the mysteries about thinking minds: how is thinking (as characterized above) possible? How is rationality mechanically possible? Then we have also seen a third argument that LOTH would partially contribute to the project of naturalizing intentionality by offering an account of how the semantic properties of whole attitudes are fixed on the basis of their atomic constituents. But there have been many other arguments for LOTH. In this section, I will describe only those arguments that have been historically more influential and controversial.

6.1 Argument from Contemporary Cognitive Psychology

When Fodor first formulated LOTH with significant elaboration in his (1975), he introduced his major argument for it along with its initial formulation in the first chapter. It was basically this: our best scientific theories and models of different aspects of higher cognition assume a framework that requires a computational/representational medium for them to be true. More specifically, he analyzed the basic form of the information processing models developed to account for three types of cognitive phenomena: *perception* as the fixation of perceptual beliefs, *concept learning* as hypothesis formation and confirmation, and *decision making* as a form of representing and evaluating the consequences of possible actions carried out in a situation with a preordered set of preferences. He rightly pointed out that all these psychological models treated mental processes as computational processes defined over representations. Then he drew what seems to be the obvious conclusion: if these models are right in at least treating mental processes as computational, even if not in detail, then there must be a LOT over which they are defined, hence LOTH.

In Fodor's (1975), the arguments for different aspects of LOTH are diffused and the emphasis, with the book's slogan “no computation without representation”, is put on the RTM rather than on (B) or (C). But all the elements are surely there.

6.2 Argument from the Productivity of Thought

People seem to be capable of entertaining an infinite number of thoughts, at least in principle, although they in fact entertain only a finite number of them. Indeed adults who speak a natural language are capable of understanding sentences they have never heard uttered before. Here is one: there is a big lake of melted gold on the dark side of the moon. I bet that you have never heard this sentence before, and yet, you have no difficulty in understanding it: it is one you in fact likely believe false. But this sentence was arbitrary, there are infinitely many such sentences I can in principle utter and you can in principle understand. But understanding a sentence is to entertain the thought/proposition it expresses. So there are in principle infinitely many thoughts you are capable of entertaining. This is sometimes expressed by saying that we have an unbounded *competence* in entertaining different thoughts, even though we have a bounded *performance*. But this unbounded capacity is to be achieved by finite means. For instance, storing an infinite number of representations in our heads is out of the question: we are finite beings. If human cognitive capacities (capacities to entertain an unbounded number of thoughts, or to have attitudes towards an unbounded number of propositions) are productive in this sense, how is this to be explained on the basis of finitary resources?

The explanation LOTH offers is straightforward: postulate a representational system that satisfies at least (B1). Indeed, recursion is the only known way to produce an infinite number of symbols from a finite base. In fact, given LOTH, productivity of thought as a competence mechanism seems to be guaranteed.[20]

6.3 Argument from the Systematicity and Compositionality of Thought

Systematicity of thought consists in the empirical fact that the ability to entertain certain thoughts is intrinsically connected to the ability to entertain certain others. Which ones? Thoughts that are related in a certain way. In what way? There is a certain initial difficulty in answering such questions. I think, partly because of this, Fodor (1987) and Fodor and Pylyshyn (1988), who are the original defenders of this kind of argument, first argue for the systematicity of language production and understanding: the ability to produce/understand certain sentences is intrinsically connected to the ability to produce/understand certain others. Given that a mature speaker is able to produce/understand a certain sentence in her native language, by psychological law, there always appear to be a cluster of other sentences that she is able to produce/understand. For instance, we don't find speakers who know how to express in their native language the fact that John loves the girl but not the fact that the girl loves John. This is apparently so, moreover, for expressions of any n-place relation.

Fodor and Pylyshyn bring out the force of this psychological fact by comparing learning languages the way we actually do with learning a language by memorizing a huge phrase book. In the phrase book model, there is nothing to prevent someone learning how to say ‘John loves the girl’ without learning how to say ‘the girl loves John.’ In fact, that is exactly the way some information booklets prepared for tourists help them to cope with their new social environment. You might, for example, learn from a phrase book how to say ‘I'd like to have a cup of coffee with sugar and milk’ in Turkish without knowing how to say/understand absolutely anything else in Turkish. In other words, the phrase book model of learning a language allows arbitrarily punctate linguistic capabilities. In contrast, a speaker's knowledge of her native language is not punctate, it is *systematic*. Accordingly, we do not find, by nomological necessity, native speakers whose linguistic capacities are punctate.

Now, how is this empirical truth (in fact, a law-like generalization) to be explained? Obviously if this is a general nomological fact, then learning one's native language cannot be modeled on the phrase book model. What is the alternative? The alternative is well known. Native speakers master the grammar and vocabulary of their language. But this is just to say that sentences are not atomic, but have syntactic constituent structure. If you have a vocabulary, the grammar tells you how to combine *systematically* the words into sentences. Hence, in this way, if you know how to construct a particular sentence out of certain words, you automatically know how to construct many others. If you view all sentences as atomic, then, as Fodor and Pylyshyn say, the systematicity of language production/understanding is a mystery, but if you acknowledge that sentences have syntactic constituent structure, systematicity of linguistic capacities is what you automatically get; it is guaranteed. This is the orthodox explanation of linguistic systematicity.

From here, according to Fodor and Pylyshyn, establishing the systematicity of thought as a nomological fact is one step away. If it is a law that the ability to understand a sentence is systematically connected to the ability to understand many others, then it is similarly a law that the ability to think a thought is systematically connected to the ability to think many others. For to understand a sentence is just to think the thought/proposition it expresses. Since, according to RTM, to think a certain thought is just to token a representation in the head that expresses the relevant proposition, the ability to token certain representations is systematically connected to the ability to token certain others. But then, this fact needs an adequate explanation too. The classical explanation LOTH offers is to postulate a system of representations with combinatorial syntax exactly as in the case of the explanation of the linguistic systematicity. This is what (B1) offers.[21] This seems to be the only explanation that does not make the systematicity of thought a miracle, and thus argues for the LOT hypothesis.

However, thought is not only systematic but also compositional: systematically connected thoughts are also always semantically related in such a way that the thoughts so related seem to be composed out of the same semantic elements. For instance, the ability to think ‘John loves the girl’ is connected to the ability to think ‘the girl loves John’ but not to, say, ‘protons are made up of quarks’ or to ‘2+2=4.’ Why is this so? The answer LOTH gives is to postulate a combinatorial semantics in addition to a combinatorial syntax, where an atomic constituent of a mental sentence makes (approximately) the same semantic contribution to any complex mental expression in which it occurs. This is what Fodor and Pylyshyn call ‘the principle of compositionality’.[22]

In brief, it is an argument for LOTH that it offers a cogent and principled solution to the systematicity and compositionality of cognitive capacities by postulating a system of representations that has a combinatorial syntax *and* semantics, i.e., a system of representations that satisfies at least (B1).

6.4 Argument from the Systematicity of Thinking (Inferential Coherence)

Systematicity of thought does not seem to be restricted solely to the systematic ability to entertain certain *thoughts*. If the system of mental representations does have a combinatorial syntax, then there is a set of rules, psychosyntactic formation rules, so to speak, that govern the construction of well-formed expressions in the system. It is this fact, (B1), that guarantees that if you can form a mental sentence on the basis of certain rules, then you can also form many others on the basis of the same rules. The rules of combinatorial syntax determine the syntactic or formal structure of complex mental representations. This is the *formative* (or, *formational*) aspect of systematicity. But inferential *thought processes* (i.e., *thinking*) seem to be systematic too: the ability to make certain inferences is intrinsically connected to the ability to make certain many others. For instance, you do not find minds that can infer '*A*' from '*A&B*' but cannot infer '*C*' from '*A&B&C*.' It seems to be a psychological fact that inferential capacities come in clusters that are homogeneous in certain aspects. How is this fact (i.e., the *inferential* or *transformational* systematicity) to be explained?

As we have seen, the explanation LOTH offers depends on the exploitation of the notion of logical form or syntactic structure determined by the combinatorial syntax postulated for the representational system. The combinatorial syntax not only gives us a criterion of well-formedness for mental expressions, but it also defines the logical form or syntactic structure for each well-formed expression. The classical solution to inferential systematicity is to make the mental operations on representations sensitive to their form or structure, i.e., to insist on (B2). Since, from a syntactic view point, similarly formed expressions will have similar forms, it is possible to define a single operation which will apply to only certain expressions that have a certain form, say, only to conjunctions, or conditionals. This allows the LOT theorist to give homogeneous explanations of what appear to be homogeneous classes of inferential capacities. This is one of the greatest virtues of LOTH, hence provides an argument for it.

The solution LOTH offers for what I called the problem of thinking, above, is connected to the argument here because the two phenomena are connected in a deep way. Thinking requires that the logico-semantic properties of a *particular* thought process be somehow causally implicated in the process (say, inferring that John is happy from knowing that if John is at the beach then John is happy and coming to realize that John is indeed at the beach). The systematicity of inferential thought processes then is based on the observation that if the agent is capable of making *that* particular inference, then she is capable of making many other somehow *similarly* organized inferences. But the idea of similar organization in this context seems to demand some sort of classification of thoughts independently of their *particular* content. But what can the basis of such a classification be? The only basis seems to be the logico-syntactic properties of thoughts, their form. Although it feels a little uneasy to talk about syntactic properties of thoughts common-sensically understood, it seems that they are forced upon us by the very attempt to understand their semantic properties: how, for instance, could we explain the semantic content of the thought that if John is at the beach then he is happy without somehow appealing to its being a *conditional*? This is the point of contact between the two phenomena. Especially when the demands of naturalism are added to this picture, inferring a LOT (= a representational system satisfying B) realized in the brain becomes almost irresistible. Indeed Rey (1995) doesn't resist and claims that, given the above observations, LOTH can be established on the basis of arguments that are not “merely empirical”. I leave it to the reader to evaluate whether mere critical reflection on our concepts of thought and thinking (along with certain mundane empirical observations about them) can be sufficient to establish LOTH.[23]

7. Objections to LOTH

There have been numerous arguments against LOTH. Some of them are directed more specifically against the Representational Theory of Mind (A), some against functionalist materialism (C). Here I will concentrate only on those arguments specifically targeting (B)—the most controversial component of LOTH.

7.1 Regress Arguments against LOTH

These arguments rely on the explanations offered by LOTH defenders for certain aspects of natural languages. In particular, many LOT theorists advert to LOTH to explain (1) how natural languages are learned, (2) how natural languages are understood, or (3) how the utterances in such languages can be meaningful. For instance, according to Fodor (1975), natural languages are learned by forming and confirming hypotheses about the translation of natural language sentences into Mentalese such as: ‘Snow is white’ is true in English if and only if *P*, where ‘*P*’ is a sentence in one's LOT. But to be able to do that, one needs a representational medium in which to form and confirm hypotheses—at least to represent the truth-conditions of natural language sentences. The LOT is such a medium. Again, natural languages are understood because, roughly, such an understanding consists in translating their sentences into one's Mentalese. Similarly, natural language utterances are meaningful in virtue of the meanings of corresponding Mentalese sentences.

The basic complaint is that in each of these cases, either the explanations generate a regress because the same sort of explanations ought to be given for how the LOT is learned, understood or can be meaningful, or else they are gratuitous because if a successful explanation can be given for LOT that does not generate a regress then it could and ought to be given for the natural language phenomena without introducing a LOT (see, e.g., Blackburn 1984). Fodor's response in (1975) is (1) that LOT is not learned, it's innate; (2) that it's understood in a different sense than the sense involved in natural language comprehension; (3) that LOT sentences acquire their meanings not in virtue of another meaningful language but in a completely different way, perhaps by standing in some sort of causal relation to what they represent or by having certain computational profiles (see above, §5.3). For many who have a Wittgensteinian bent, these replies are not likely to be convincing. But here the issues tend to concern RTM rather than (B).

Laurence and Margolis (1997) point out that the regress arguments depend on the assumption that LOTH is introduced only to explain (1)-(3). If it can be shown that there are lots of other empirical phenomena for which the LOTH provides good explanations, then the regress arguments fail because LOTH then would not be gratuitous. In fact, as we have seen above, there are plenty of such phenomena. But still it is important to realize that the sort of explanations proposed for the understanding of one's LOT (computational use/activity of LOT sentences with certain meanings) and how LOT sentences can be meaningful (computational roles and/or nomic relations with the world) cannot be given for (1)-(3): it's unclear, for example, what it would be like to give a computational role and/or nomic relation account for the meanings of natural language utterances. (See Knowles 1998 for a reply to Laurence & Margolis 1997; Margolis & Laurence 1998 counterreplies to Knowles.)

7.2 Propositional Attitudes without Explicit Representations

Dennett in his review of Fodor's (1975) has raised the following objection (cf. Fodor 1987: 21–3 for a similar discussion):

> In a recent conversation with the designer of a chess-playing program I heard the following criticism of a rival program: “it thinks it should get its queen out early.” This ascribes a propositional attitude to the program in a very useful and predictive way, for as the designer went on to say, one can usefully count on chasing that queen around the board. But for all the many levels of explicit representation to be found in that program, nowhere is anything roughly synonymous with “I should get my queen out early” explicitly tokened. The level of analysis to which the designer's remark belongs describes features of the program that are, in an entirely innocent way, emergent properties of the computational processes that have “engineering reality.” I see no reason to believe that the relation between belief-talk and psychological talk will be any more direct. (Dennett 1981: 107)

The objection, as Fodor (1987: 22) points out, isn't that the program has a *dispositional*, or *potential*, belief that it will get its queen out early. Rather, the program actually operates on this belief. There appear to be lots of other examples: e.g., in reasoning we pretty often follow certain inference rules like modus ponens, disjunctive syllogism, etc., without necessarily explicitly representing them.

The standard reply to such objections is to draw a distinction between rules on the basis of which Mentalese data-structures are manipulated, and the data-structures themselves (intuitively, the program/data distinction). LOTH is not committed to every rule's being explicitly represented. In fact, as a point of nomological fact, in a computational device not every rule can be explicitly represented: some *have to* be hard-wired and, thus, implicit in this sense. In other words, LOTH permits but doesn't require that rules be explicitly represented. On the other hand, data structures *have to* be explicitly represented: it is these that are manipulated formally by the rules. No causal manipulation is possible without explicit tokening of these structures. According to Fodor, if a propositional attitude is an actual episode in one's reasoning that plays a causal role, then LOTH is committed to explicit representation of its content, which is as per (A2 and B2) causally implicated in the physical process realizing that reasoning. Dispositional propositional attitudes can then be accounted for in terms of an appropriate principle of inferential closure of explicitly represented propositional attitudes (cf. Lycan 1986).

Dennett's chess program certainly involves explicit representations of the chess board, the pieces, etc. and perhaps some of the rules. Which rules are implicit and which are explicit depend on the empirical details of the program. Pointing to the fact that there may be some rules that are emergent out of the implementation of explicit rules and data-structures does not suffice to undermine LOTH.

7.3 Explicit Representations without Propositional Attitudes

In any sufficiently complex computational system, there are bound to be many symbol manipulations with no obviously corresponding description at the level of propositional attitudes. For instance, when a multiplication program is run through a standard conventional computer, the steps of the program are translated into the computer's machine language and executed there, but at this level the operations apply to 1's and 0's with no obvious way to map them onto the original numbers to be multiplied or to the multiplication operation. So it seems that at those levels that, according to Dennett, have engineering reality there are plenty of explicit tokenings of symbols with appropriate operations over them that don't correspond to anything like the propositional attitudes of folk psychology. In other words, there is plenty of symbolic activity which it would be wrong to say a *person* engages in. Rather, they are done by the person's subpersonal computational *components* as opposed to the person. How to rule out such cases? (cf. Fodor 1987: 23–6 for a similar discussion.)

They are ruled out by an appropriate reading of (A1) and (B1): (A1) says that the person herself must stand in an appropriate computational relation to a Mentalese sentence, which, as per (B1), has a suitable syntax and semantics. Only then will the sentence constitute the person's having a propositional attitude. Not all explicit symbols in one's LOT will satisfy this. In other words, not every computational routine will correspond to a processing appropriately described as storage in, e.g., the “belief-box”. Furthermore, as pointed out by Fodor (1987), LOTH would vindicate the common sense view of propositional attitudes if they turn out to be computational relations to Mentalese sentences. It may not be further required that every explicit representation correspond to a propositional attitude.

There have been many other objections to LOTH in recent years raised especially by connectionists: that LOT systems cannot handle certain cognitive tasks like perceptual pattern recognition, that they are too brittle and not sufficiently damage resistant, that they don't exhibit graceful degradation when physically damaged or as a response to noisy or degraded input, that they are too rigid, deterministic, so are not well-suited for modeling humans' capacity to satisfy multiple soft-constraints so gracefully, that they are not biologically realistic, and so on. (For useful discussions of these and many similar objections, see Rumelhart, McClelland and the PDP Research Group (1986), Fodor and Pylyshyn (1988), Horgan and Tienson (1996), Horgan (1997), McLaughlin and Warfield (1994), Bechtel and Abrahamsen (2002), Marcus (2002).)

8. The Connectionism/Classicism Debate

When Jerry Fodor published his influential book, *The Language of Thought*, in (1975), he called LOTH “the only game in town.” As we have seen, it was the philosophical articulation of the assumptions that underlay the new developments in “cognitive sciences” after the demise of behaviorism. Fodor argued for the truth of LOTH on the basis of the successes of the best scientific theories we had then. Indeed most of the scientific work in cognitive psychology, psycholinguistics, and AI assumed the framework of LOTH.

In the early 1980's, however, Fodor's claim that LOTH was the only game in town was beginning to be challenged by some who were working on so-called connectionist networks. They claimed that *connectionism* offered a new and radically different alternative to classicism in modeling cognitive phenomena. The name ‘classicism’ has since then become to be applied to the LOTH framework. On the other hand, many classicists like Fodor thought that connectionism was nothing but a slightly more sophisticated way with which the old and long dead associationism, whose roots could be traced back to early British empiricists, was being revived. In 1988 Fodor and Pylyshyn (F&P) published a long article, “Connectionism and Cognitive Architecture: A Critical Analysis”, in which they launched a formidable attack on connectionism, which largely set the terms for the ensuing debate between connectionists and classicists.

F&P's forceful criticism consists in posing a dilemma for connectionists: They either fail to explain the law-like cognitive regularities like systematicity and productivity in an adequate way or the connectionist models are nothing but mere implementation models of classical architectures; hence, they fail to provide a radically new paradigm as connectionists claim. This conclusion was also meant to be a challenge: Explain the cognitive regularities in question without postulating a LOT architecture.

First, let me present F&P's argument against connectionism in a somewhat reconstructed fashion. It will be helpful to characterize the debate by locating the issues according to the reactions many connectionists had to the premises of the argument.

F&P's Argument against Connectionism in their (1988) article:

i. Cognition essentially involves representational states and causal operations whose domain and range are these states; consequently, any scientifically adequate account of cognition should acknowledge such states and processes.

ii. Higher cognition (specifically, thought and thinking with propositional content) conceived in this way, has certain empirically interesting properties: in particular, it is a law of nature that cognitive capacities are *productive*, *systematic*, and *inferentially coherent*.

iii. Accordingly, the architecture of any proposed cognitive model is scientifically adequate only if it guarantees that cognitive capacities are productive, systematic, etc. This would amount to explaining, in the scientifically relevant and required sense, how it could be a law that cognition has these properties.

iv. The only way (i.e., necessary condition) for a cognitive architecture to guarantee systematicity (etc.) is for it to involve a representational system for which (B) is true (see above). (Classical architectures necessarily satisfy (B).)

v. Either the architecture of connectionist models does satisfy (B), or it does not.

vi. If it does, then connectionist models are implementations of the classical LOT architecture and have little new to offer (i.e., they fail to compete with classicism, and thus connectionism does not constitute a radically new way of modeling cognition).

vii. If it does not, then (since connectionism does not then guarantee systematicity, etc., in the required sense) connectionism is empirically false as a theory of the *cognitive* architecture.

viii. Therefore, connectionism is either true as an implementation theory, or empirically false as a theory of cognitive architecture.

The notion of *cognitive architecture* assumes special importance in this debate. F&P's characterization of the notion goes as follows:

> The architecture of the cognitive system consists of the set of basic operations, resources, functions, principles, etc. (generally the sorts of properties that would be described in a "user's manual" for that architecture if it were available on a computer) whose domain and range are the *representational states* of the organism. (1988: 10)

Also, note that (B1) and (B2) are meta-architectural properties in that they are themselves conditions upon any *specific* architecture's being classical. They define classicism per se, but not any particular way of being classical. Classicism as such simply claims that whatever the *particular* cognitive architecture of the brain might turn out to be (whatever the *specific* grammar of Mentalese turns out to be), (B) must be true of it. F&P claim that this is the only way an architecture can be said to guarantee the nomological necessity of cognitive regularities like systematicity, etc. This seems to be the relevant and required sense in which a scientific explanation of cognition is required to guarantee the regularities—hence the third premise in their argument.

Connectionist responses have fallen into four classes:

1. *Deny premise(i)*. The rejection of (i) commits connectionists to what is sometimes called *radical* or *eliminativist connectionism*. Premise (i), as F&P point out, draws a general line between eliminativism and representationalism (or, intentional realism). There has been some controversy as to whether connectionism constitutes a serious challenge to the fundamental tenets of folk psychology.[24] Although it may still be too early for assessment,[25] the connectionist research program has been overwhelmingly cognitivist: most connectionists do in fact advance their models as having causally efficacious representational states, and explicitly endorse F&P's first premise. So they seem to accept intentional realism.[26]

2. *Accept the conclusion*. This group may be seen as more or less accepting the cogency of the entire argument, and characterizes itself as *implementationalist*: they hold that connectionist networks will *implement* a classical architecture or language of thought. According to this group, the appropriate niche for neural networks is closer to neuroscience than to cognitive psychology. They seem to view the importance of the program in terms of its prospects of closing the gap between the neurosciences and high-level cognitive theorizing. In this, many seem content to admit premise (vi). (See Marcus 2001 for a discussion of the virtues of placing connectionist models closer to implementational level.)

3. *Deny premise (ii) or (iv)*. Some connectionists reject (ii) or (iv),[27] holding that there are no lawlike cognitive regularities such as systematicity (etc.) to be explained, or that such regularities do not require a (B)-like architecture for their explanation.

Those who question (ii) often question the empirical evidence for systematicity (etc.) and tend to ignore the challenge put forward by F&P. Those who question (iv) also often question (ii), or they argue that there can be very different sort of explanations for systematicity and the like (e.g. evolutionary explanations, see Braddon-Mitchell and Fitzpatrick 1990), or they question the very notion of explanation involved (e.g. Matthews 1994). There are indeed quite a number of different kinds of arguments in the literature against these premises.[28] For a sampling, see Aydede (1995) and McLaughlin (1993b), who partitions the debate similarly.

4. *Deny premise (vi)*. The group of connectionists who have taken F&P's challenge most seriously has tended to reject premise (vi) in their argument, while accepting, on the face of it, the previous five premises (sometimes with reservations on the issue of productivity). They think that it is possible for connectionist representations to be syntactically structured in some sense without being classical. Prominent in this group are Smolensky (1990a, 1990b, 1995), van Gelder (1989, 1990, 1991), Chalmers (1990, 1993).[29] Some connectionists whose models give support to this line include Elman (1989), Hinton (1990), Touretzky (1990), Pollack (1990), Barnden and Srinivas (1991), Shastri and Ajjanagadde (1993), Plate (1998), Hummel et al. (2004), Van Der Velde and De Kamps (2006), Barrett et al. (2008), Sanjeevi and Bhattacharyya (2010).

Much of the recent debate between connectionists and classicists has focused on this option. How is it possible to reject premise (vi), which seems true by definition of classicism. The connectionists' answer, roughly put, is that when you devise a representational system whose satisfaction of (B) relies on a *non-concatenative* realization of structural/syntactic complexity of representations, you have a non-classical system. (See especially Smolensky 1990a and van Gelder 1990.) Interestingly, some classicists like Fodor and McLaughlin (1990) (F&M) seem to agree. F&M stipulate that you have a classical system only if the syntactic complexity of representations is realized *concatenatively*, or as it is sometimes put, *explicitly*:

> We … stipulate that for a pair of expression types E1, E2, the first is a *Classical* constituent of the second *only if* the first is tokened whenever the second is tokened. (F&M 1990: 186)

The issues about how connectionists propose to obtain constituent structure non-concatenatively tend to be complex and technical. But they propose to exploit so called *distributed representations* in certain novel ways. The essential idea behind most of them is to use vector (and tensor) algebra (involving superimposition, multiplication, etc. of vectors) in composing and decomposing connectionist representations which consist in coding patterns of activity across neuron-like units which can be modeled as vectors. The result of such techniques is the production of representations that have in some interesting sense a complexity whose constituent structure is largely implicit in that the constituents are not tokened explicitly when the representations are tokened, but can be recovered by further operations upon them. The interested reader should consult some of the pioneering work by Elman (1989), Hinton (1990), Smolensky (1989, 1990, 1995), Touretzky (1990), Pollack (1990).

F&M's criticism, more specifically stated, however, is this. Connectionists with such techniques only satisfy (B1) in some “extended sense”, but they are incapable of satisfying (B2), precisely because their way of satisfying (B1) is committed to a non-concatenative realization of syntactic structures.

Some connectionists disagree (e.g., Chalmers 1993, Niklasson and van Gelder 1994—see also Browne 1998 and Browne and Sun 2001 for discussion and overview of models): they claim that you can have structure-sensitive transformations or operations defined over representations whose syntactic structure is non-concatenatively realized. So given the apparent agreement that non-concatenative realization is what makes a system non-classical, connectionists claim that they can and do perfectly satisfy (B) in its entirety with their connectionist models without implementing classical models.

The debate still continues and there is a growing literature built around the many issues raised by it. Aydede (1997a) offers an extensive analysis of the debate between classicists and this group of connectionists with special attention to the conceptual underpinnings of the debate. (See also Roth 2005 who argues that to the extent to which connectionist models can transform representations successfully according to an algorithmic function, to that extent they count as executing program in the sense relevant to classical program execution.) Aydede argues that both parties are wrong in assuming that concatenative realization is relevant to the characterization of LOTH. Part of the argument is that concatenative realization of (B) is just that—a realization. The attentive reader might have noticed that there is nothing in the characterization of (B) that requires concatenative realization. Indeed, when we look at all the major arguments for LOTH focused on the need for (B), none of them requires concatenation or explicit realization of syntactic structure. In fact, it is almost on the border of confusion to necessarily associate LOTH to such an implementational level issue. If anything, this class of connectionist networks, if successful and generalizable across all higher cognition, contributes to our understanding of how radically differently a LOTH architecture could be implemented in neural networks. Indeed, if these models prove to be adequate for explaining the full range of human cognitive capacities, they would show how syntactically structured representations and structure sensitive processes could be implemented in a radically new way. So research programs in this niche are by no means trivial or insignificant. But we need to be clear and careful about what minimally needs to be the case for LOTH to be true, and why.

On the other hand, it is by no means clear that these connectionist models are successful and generalizable (scalable). They all have proved to have serious limitations that seem to be tied to their particular ways of implementing variable binding (syntactic structure) and structure sensitive processing. For critical discussion, see Marcus (2001), Hadley (2009), Browne and Sun (2001). Marcus in particular makes a strong and largely empirical case for why classical symbol systems are needed for explaining human capacities of variable binding and generalizing, and why existing connectionist models aren't up to the job to match human capacities while remaining non-classical. Indeed the trend in the last fifteen years seems to be towards developing hybrid systems combining connectionist and classical symbol processing models—see, for instance, the articles in Wermter and Sun (2000).[30]

Bibliography

- Aizawa, K. (1994). “Representations without Rules, Connectionism and the Syntactic Argument.” *Synthese* 101(3): 465–492.
- –—. (1997a). “Explaining Systematicity.” *Mind and Language* 12(2): 115–136.
- –—. (1997b). “Exhibiting versus Explaining Systematicity: A Reply to Hadley and Hayward.” *Minds and Machines* 7(1): 39–55.
- –—. (2003). *The Systematicity Arguments*, Kluwer Academic Publishers.
- Aydede, Murat. (1995). “Connectionism and Language of Thought”, *CSLI Technical Report*, Stanford, CSLI, 95–195. (This is an early version of Aydede 1997 but contains quite a lot of expository material not contained in 1997.)
- –—. (1997a). “Language of Thought: The Connectionist Contribution,” *Minds and Machines*, Vol. 7, No. 1, pp. 57–101.
- –—. (1997b). “Has Fodor Really Changed His Mind on Narrow Content?”, *Mind and Language*, 12(3–4): 422–458.
- –—. (1998). “Fodor On Concepts and Frege Puzzles,” *Pacific Philosophical Quarterly*, 79(4): 289–294.
- –—. (2000). “On the Type/Token Relation of Mental Representations,” *Facta Philosophica: International Journal for Contemporary Philosophy*, 2(1): 23–49.
- –—. (2005). “Computation and Functionalism: Syntactic Theory of Mind Revisited” in Gürol Irzik and G. Güzeldere (eds.), *Boston Studies in the History and Philosophy of Science*, Dordrecht: Kluwer Academic Publishers.
- Aydede, Murat, and Güven Güzeldere (2005). “Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness”, *Noûs*, 39(2): 197–255.
- Armstrong, D.M. (1973). *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.

- –––. (1980). *The Nature of Mind*, Ithaca, NY: Cornell University Press.
- Bader, S. and B. Hitzler (2005). “Dimensions of neural-symbolic integration—a structured survey” in *We Will Show Them: Essays in Honour of Dov Gabbay*, edited by S. Artemov and H. Barringer and A. S. d'Avila Garcez and L.C. Lamb and J. Woods, King's College Publications.
- Barnden, J. and K. Srinivas (1991). “Encoding techniques for complex information structures in connectionist systems,” *Connection Science,* 3(3): 269–315.
- Barrett, L., J Feldman, and L. Mac Dermed (2008). “A (somewhat) new solution to the variable binding problem,” *Neural Computation*, Vol. 20, pp. 2361–2378.
- Barsalou, L. W. (1993). “Flexibility, Structure, and Linguistic Vagary in Concepts: Manifestations of a Compositional System of Perceptual Symbols” in *Theories of Memory*, edited by A. Collins, S. Gathercole, M. Conway and P. Morris, Hillsdale, NJ: Lawrence Erlbaum Associates.
- –––. (1999). “Perceptual Symbol Systems.” *Behavioral and Brain Sciences* 22(4).
- Barsalou, L. W., W. Yeh, B. J. Luka, K. L. Olseth, K. S. Mix, and L.-L. Wu. (1993). “Concepts and Meaning”, *Chicago Linguistics Society* 29.
- Barsalou, L. W., and J. J. Prinz. (1997). “Mundane Creativity in Perceptual Symbol Systems” in *Creative Thought: An Investigation of Conceptual Structures and Processes*, edited by T. B. Ward, S. M. Smith and J. Vaid, Washington, DC: American Psychological Association.
- Barwise, Jon and John Etchemendy (1995). *Hyperproof*, Stanford, Palo Alto: CSLI Publications.
- Barwise, J. and J. Perry (1983). *Situations and Attitudes*, Cambridge, Massachusetts: MIT Press.
- Bechtel, W. and A. Abrahamsen (2002). *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*, 2nd Edition, Oxford, UK: Basil Blackwell.
- Blackburn, S. (1984). *Spreading the Word*, Oxford, UK: Oxford University Press.
- Block, Ned. (1980). “Troubles with Functionalism” in *Readings in Philosophy of Psychology*, N. Block (ed.), Vol.1, Cambridge, Massachusetts: Harvard University Press, 1980. (Originally appeared in *Perception and Cognition: Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science*, C.W. Savage (ed.), Minneapolis: The University of Minnesota Press, 1978.)
- –––. (ed.) (1981). *Imagery*. Cambridge, Massachusetts: MIT Press.
- –––. (1983a). “Mental Pictures and Cognitive Science,” *Philosophical Review* 93: 499–542. (Reprinted in *Mind and Cognition*, W.G. Lycan (ed.), Oxford, UK: Basil Blackwell, 1990.)
- –––. (1983b). “The Photographic Fallacy in the Debate about Mental Imagery”, *Nous* 17: 651–62.
- ––– (1986). “Advertisement for a Semantics for Psychology” in *Studies in the*

Philosophy of Mind: Midwest Studies in Philosophy, Vol.10, P. French, T. Euhling and H. Wettstein (eds.), Minneapolis: University of Minnesota Press.

- Braddon-Mitchell, David and John Fitzpatrick (1990). “Explanation and the Language of Thought,” *Synthese* 83: 3–29.
- Braddon-Mitchell, D. and F. Jackson (2007). *Philosophy of Mind and Cognition: An Introduction*, Blackwell.
- Browne, A. (1998). “Performing a symbolic inference step on distributed representations”, *Neurocomputing*, 19(1–3): 23–34.
- Browne, A., and R. Sun (1999). “Connectionist variable binding”, *Expert Systems*, 16(3): 189–207.
- –––. (2001). “Connectionist inference models”, *Neural Networks*, 14(10): 1331–1355.
- Brentano, Franz (1874/1973). *Psychology from an Empirical Standpoint*, A. Rancurello, D. Terrell and L. McAlister (trans.), London: Routledge and Kegan Paul.
- Butler, Keith (1991). “Towards a Connectionist Cognitive Architecture,” *Mind and Language*, Vol. 6, No. 3, pp. 252–72.
- Chalmers, David J. (1990). “Syntactic Transformations on Distributed Representations,” *Connection Science*, Vol. 2.
- –––. (1993). “Connectionism and Copositionality: Why Fodor and Pylyshyn Were Wrong” in *Philosophical Psychology* 6: 305–319.
- –––. (1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford, UK: Oxford University Press.
- Churchland, Patricia Smith (1986). *Neurophilosophy: Toward a Unified Science of Mind-Brain*, Cambridge, Massachusetts: MIT Press.
- –––. (1987). “Epistemology in the Age of Neuroscience,” *Journal of Philosophy*, Vol. 84, No. 10, pp. 544–553.
- Churchland, Patricia S. and Terrence J. Sejnowski (1989). “Neural Representation and Neural Computation” in *Neural Connections, Neural Computation*, L. Nadel, L.A. Cooper, P. Culicover and R.M. Harnish (eds.), Cambridge, Massachusetts: MIT Press, 1989.
- Churchland, Paul M. (1990). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, Massachusetts: MIT Press.
- –––. (1981). “Eliminative Materialism and the Propositional Attitudes,” *Journal of Philosophy* 78: 67–90.
- Churchland, Paul M. and P.S. Churchland (1990). “Could a Machine Think?,” *Scientific American*, Vol. 262, No. 1, pp. 32–37.
- Clark, Andy (1988). “Thoughts, Sentences and Cognitive Science,” *Philosophical Psychology*, Vol. 1, No. 3, pp. 263–278.
- –––. (1989a). “Beyond Eliminativism,” *Mind and Language*, Vol. 4, No. 4, pp. 251–279.

- –—. (1989b). *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*, Cambridge, Massachusetts: MIT Press.
- –—. (1990). "Connectionism, Competence, and Explanation," *British Journal for Philosophy of Science*, 41: 195–222.
- –—. (1991). "Systematicity, Structured Representations and Cognitive Architecture: A Reply to Fodor and Pylyshyn" in *Connectionism and the Philosophy of Mind*, Terence Horgan and John Tienson (eds.), *Studies in Cognitive Systems* (Volume 9), Dordrecht: Kluwer Academic Publishers, 1991.
- –—. (1994). "Language of Thought (2)" in *A Companion to the Philosophy of Mind* edited by S. Guttenplan, Oxford, UK: Basil Blackwell, 1994.
- Cowie, F. (1998). *What's Within? Nativism Reconsidered*. Oxford, UK, Oxford University Press.
- Cummins, Robert. (1986). "Inexplicit Information" in The Representation of Knowledge and Belief, M. Brand and R.M. Harnish (eds.), Tucson, Arizona: Arizona University Press, 1986.
- –—. (1989). *Meaning and Mental Representation*, Cambridge, Massachusetts: MIT Press.
- –—. (1996). *Representations, Targets, and Attitudes*, Cambridge, Massachusetts: MIT Press.
- Cummins, Robert and Georg Schwarz (1987). "Radical Connectionism," *The Southern Journal of Philosophy*, Vol. XXVI, Supplement.
- Davidson, Donald (1984). *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.
- Davies, Martin (1989). "Connectionism, Modularity, and Tacit Knowledge," *British Journal for the Philosophy of Science* 40: 541–555.
- –—. (1991). "Concepts, Connectionism, and the Language of Thought," in *Philosophy and Connectionist Theory*, W. Ramsey, S.P. Stich and D.E. Rumelhart (eds.), Hillsdale, NJ: Lawrence Erlbaum, 1991.
- –—. (1995). "Two Notions of Implicit Rules," *Philosophical Perspectives* 9: 153–83.
- Dennett, D.C. (1978). "Two Approaches to Mental Images" in *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, Massachusetts: MIT Press, 1981.
- –—. (1981). "Cure for the Common Code" in *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, Massachusetts: MIT Press, 1981. (Originally appeared in *Mind*, April 1977.)
- –—. (1986). "The Logical Geography of Computational Approaches: A View from the East Pole" in *The Representation of Knowledge and Belief*, Myles Brand and Robert M. Harnish (eds.), Tucson: The University of Arizona Press, 1986.
- –—. (1991a). "Real Patterns," *Journal of Philosophy*, Vol. LXXXVIII, No. 1, pp. 27–51.

- —–. (1991b). “Mother Nature Versus the Walking Encyclopedia: A Western Drama” in *Philosophy and Connectionist Theory*, W. Ramsey, S.P. Stich and D.E. Rumelhart (eds.), Lawrence Erlbaum Associates.
- Descartes, R. (1637/1970). “Discourse on the Method” in *The Philosophical Works of Descartes*, Vol.I, E.S. Haldane and G.R.T. Ross (trans.), Cambridge, UK: Cambridge University Press.
- Devitt, Michael (1990). “A Narrow Representational Theory of the Mind,” *Mind and Cognition*, W.G. Lycan (ed.), Oxford, UK: Basil Blackwell, 1990.
- —–. (1996). *Coming to our Senses: A Naturalistic Program for Semantic Localism*, Cambridge, UK: Cambridge University Press.
- Devitt, Michael and Sterelny, Kim (1987). *Language and Reality: An Introduction to the Philosophy of Language*, Cambridge, Massachusetts: MIT Press.
- Dretske, Fred (1981). *Knowledge and the Flow of Information*, Cambridge, Massachusetts: MIT Press.
- —–. (1988). *Explaining Behavior*, Cambridge, Massachusetts: MIT Press.
- Elman, Jeffrey L. (1989). “Structured Representations and Connectionist Models”, *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society*, Ann Arbor, Michigan, pp.17–23.
- Field, Hartry H. (1972). “Tarski's Theory of Truth”, *Journal of Philosophy,* 69: 347–75.
- —–. (1978). “Mental Representation”, *Erkenntnis* 13, 1, pp.9–61. (Also in *Mental Representation: A Reader*, S.P. Stich and T.A. Warfield (eds.), Oxford, UK: Basil Blackwell, 1994. References in the text are to this edition.)
- Fodor, Jerry A. (1975). *The Language of Thought*, Cambridge, Massachusetts: Harvard University Press.
- —–. (1978). “Propositional Attitudes” in *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, J.A. Fodor, Cambridge, Massachusetts: MIT Press, 1981. (Originally appeared in *The Monist* 64, No.4, 1978.)
- —–. (1978a). “Computation and Reduction” in *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, J.A. Fodor, Cambridge, MA: MIT Press. (Originally appeared in *Minnesota Studies in the Philosophy of Science: Perception and Cognition*, Vol. 9, W. Savage (ed.), 1978.)
- —–. (1978b). “Tom Swift and His Procedural Grandmother,” *Cognition*, Vol. 6. (Also in *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, J.A. Fodor, Cambridge, Massachusetts: MIT Press, 1981.)
- —–. (1980). “Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology”, *Behavioral and Brain Sciences* 3, 1, 1980. (Also in *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, J.A. Fodor, Cambridge, MA: MIT Press, 1981. References in the text are to this edition.)

- –––. (1981a). *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge, Massachusetts: MIT Press.
- –––. (1981b), “Introduction: Something on the State of the Art” in *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, J.A. Fodor, Cambridge, Massachusetts: MIT Press, 1981.
- –––. (1983). *The Modularity of Mind*, Cambridge, Massachusetts: MIT Press.
- –––. (1985). “Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum”, *Mind* 94, 1985, pp.76–100. (Also in *A Theory of Content and Other Essays*, J.A. Fodor, Cambridge, Massachusetts: MIT Press. References in the text are to this edition.)
- –––. (1986). “Banish DisContent” in *Language, Mind, and Logic*, J. Butterfield (ed.), Cambridge, UK: Cambridge University Press, 1986. (Also in *Mind and Cognition*, William Lycan (ed.), Oxford, UK: Basil Blackwell, 1990.)
- –––. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, Massachusetts: MIT Press.
- –––. (1989). “Substitution Arguments and the Individuation of Belief” in *A Theory of Content and Other Essays*, J. Fodor, Cambridge, Massachusetts: MIT Press, 1990. (Originally appeared in *Method, Reason and Language*, G. Boolos (ed.), Cambridge, UK: Cambridge University Press, 1989.)
- –––. (1990). *A Theory of Content and Other Essays*, Cambridge, Massachusetts: MIT Press.
- –––. (1991). “Replies” (Ch.15) in *Meaning in Mind: Fodor and his Critics*, B. Loewer and G. Rey (eds.), Oxford, UK: Basil Blackwell, 1991.
- –––. (2001). “Doing without What's Within: Fiona Cowie's Critique of Nativism.” *Mind*: 110(437) 99–148.
- –––. (2008). *LOT 2: The Language of Thought Revisited*, Oxford: Oxford University Press.
- Fodor, Jerry A. and Ernest Lepore (1991). “Why Meaning (Probably) Isn't Conceptual Role?”, *Mind and Language*, Vol. 6, No. 4, pp. 328–43.
- Fodor, Jerry A. and B. McLaughlin (1990). “Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work,” *Cognition* 35: 183–204.
- Fodor, Jerry A. and Zenon W. Pylyshyn (1988). “Connectionism and Cognitive Architecture: A Critical Analysis” in S. Pinker and J. Mehler, eds., *Connections and Symbols*, Cambridge, Massachusetts: MIT Press (A *Cognition* Special Issue).
- Grice, H.P. (1957). “Meaning”, *Philosophical Review*, 66: 377–88.
- Hadley, R. F. (1995). “The ”Explicit-Implicit“ Distinction.” *Minds and Machines* 5(2): 219–242.
- –––. (1997). “Cognition, Systematicity and Nomic Necessity.” *Mind and Language* 12(2): 137–153.
- –––. (1997). “Explaining Systematicity: A Reply to Kenneth Aizawa.” *Minds and*

Machines 7(4): 571–579.

- ––—. (1999). “Connectionism and Novel Combinations of Skills: Implications for Cognitive Architecture.” *Minds and Machines* 9(2): 197–221.
- ––—. (2009). “The problem of rapid variable creation,” *Neural Computation*, 21: 510–32.
- Hadley, R. F. and M. B. Hayward (1997). “Strong Semantic Systematicity from Hebbian Connectionist Learning.” *Minds and Machines* 7(1): 1–37.
- Harman, Gilbert (1973). *Thought*, Princeton University Press.
- Haugeland, John (1981). “The Nature and Plausibility of Cognitivism,” *Behavioral and Brain Sciences* I, 2: 215–60 (with peer commentary and replies).
- ––—. (1985). *Artificial Intelligence: The Very Idea*, Cambridge, Massachusetts: MIT Press.
- Hinton, Geoffrey (1990). “Mapping Part-Whole Hierarchies into Connectionist Networks,” *Artificial Intelligence*, Vol. 46, Nos. 1–2, (Special Issue on Connectionist Symbol Processing).
- Horgan, T. E. and J. Tienson (1996). *Connectionism and the Philosophy of Psychology*, Cambridge, Massachusetts: MIT Press.
- Horgan, T. (1997). “Connectionism and the Philosophical Foundations of Cognitive Science.” *Metaphilosophy* 28(1–2): 1–30.
- Hummel, J. E., Holyoak, K. J., Green, C., Doumas, L. A. A., Devnich, D., Kittur, A., & Kalar, D.J. (2004). A Solution to the Binding Problem for Compositional Connectionism. In S.D. Levy & R. Gayler: *Compositional Connectionism in Cognitive Science: Papers from the AAAI Fall Symposium* (pp. 31–34). Menlo Park, CA: AAAI Press.
- Jacob, P. (1997). *What Minds Can Do: Intentionality in a Non-Intentional World*. Cambridge, UK, Cambridge University Press.
- Kirsh, D. (1990). “When Is Information Explicitly Represented?” in *Information, Language and Cognition*. P. Hanson (ed.), University of British Columbia Press.
- Knowles, J. (1998). “The Language of Thought and Natural Language Understanding.” *Analysis* 58(4): 264–272.
- Kosslyn, S.M. (1980). *Image and Mind*. Cambridge, Massachusetts: Harvard University Press.
- ––—. (1981). “The Medium and the Message in Mental Imagery: A Theory” in *Imagery,* N. Block (ed.), Cambridge, Massachusetts: MIT Press, 1981.
- ––—. (1994). *Image and Brain*, Cambridge, Massachusetts: MIT Press.
- Kulvicki, J. (2004). “Isomorphism in information-carrying systems”, *Pacific Philosophical Quarterly* 85(4): 380–395.
- ––—. (2006). *On Images: Their Structure and Content*, Oxford: Clarendon Press.
- Laurence, Stephen and Eric Margolis (1997). “Regress Arguments Against the Language of Thought”, *Analysis*, Vol. 57, No. 1.

- –––. (2002). “Radical Concept Nativism.” *Cognition* 86: 22–55.
- Leeds, S. (2002). “Perception, Transparency, and the Language of Thought.” *Noûs* 36(1): 104–129.
- Lewis, David (1972). “Psychophysical and Theoretical Identifications,” *Australasian Journal of Philosophy*, 50(3):249–58. (Also in *Readings in Philosophy of Psychology*, Ned Block (ed.), Vols.1, Cambridge, Massachusetts: Harvard University Press, 1980.)
- –––. (1994). “Reduction of Mind” in *A Companion to the Philosophy of Mind*, edited by Samuel Guttenplan, Oxford: Blackwell, pp. 412–31.
- Loar, Brian F. (1982a). *Mind and Meaning*, Cambridge, UK: Cambridge University Press.
- –––. (1982b). “Must Beliefs Be Sentences?” in *Proceedings of the Philosophy of Science Association for 1982*, Asquith, P. and T. Nickles (eds.), East Lansing, Michigan, 1983.
- Lycan, William G. (1981). “Toward a Homuncular Theory of Believing,” *Cognition and Brain Theory* 4(2): 139–159.
- –––. (1986). “Tacit Belief” in *Belief: Form, Content, and Function,* R. Bogdan (ed.), Oxford, UK: Oxford University Press.
- –––. (1993). “A Deductive Argument for the Representational Theory of Thinking,” *Mind and Language*, Vol. 8, No. 3, pp. 404–22.
- –––. (1997). “Consciousness as Internal Monitoring” in *The Nature of Consciousness: Philosophical Debates*, edited by N. Block, O. Flanagan and G. Güzeldere, Cambridge, Massachusetts: MIT Press.
- Marcus, G. F. (1998). “Can connectionism save constructivism?” *Cognition* 66: 153–182.
- –––. (1998). “Rethinking Eliminative Connectionism.” *Cognitive Psychology* 37: 243–282.
- –––. (2001). *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. Cambridge, MA, MIT Press.
- Margolis, Eric (1998). “How to Acquire a Concept?”, *Mind and Language*.
- Margolis, E. and S. Laurence (1999). “Where the Regress Argument Still Goes Wrong: Reply to Knowles.” *Analysis* 59(4): 321–327.
- –––. (2001). “The Poverty of the Stimulus Argument.” *British Journal for the Philosophy of Science* 52: 217–276.
- ––– (forthcoming-a). “Learning Matters: The Role of Learning in Concept Acquisition.”
- –––. (forthcoming-b). “The Nativist Manifesto.”
- Markic, O. (2001). “Is Language of Thought a Conceptual Necessity?” *Acta Analytica* 16(26): 53–60.
- Marr, David (1982). *Vision*, San Francisco: W. H. Freeman.

- Martinez, F. and J. Ezquerro Martinez (1998). “Explicitness with Psychological Ground.” *Minds and Machines* 8(3): 353–374.
- Matthew, Robert J. (1994). “Three-Concept Monte: Explanation, Implementation and Systematicity”, *Synthese*, Vol. 101, No. 3, pp. 347–63.
- McGinn, Colin (1989). *Mental Content*, Oxford: Blackwell.
- –––. (1991). *The Problem of Consciousness*, Oxford, UK: Basil Blackwell.
- McLaughlin, B.P. (1993a). “The Connectionism/Classicism Battle to Win Souls,” *Philosophical Studies* 71: 163–90.
- –––. (1993b). “Systematicity, Conceptual Truth, and Evolution,” in *Philosophy and Cognitive Science,* C. Hookway and D. Peterson (eds.), Royal Institute of Philosophy, Supplement No. 34.
- McLaughlin, B.P. and Ted Warfield (1994). “The Allures of Connectionism Reexamined”, *Synthese* 101, pp. 365–400
- Millikan, Ruth Garrett (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*, Cambridge, Massachusetts: MIT Press.
- –––. (1993). *White Queen Psychology and Other Essays for Alice*, Cambridge, Massachusetts: MIT Press.
- Niklasson, L. and T. van Gelder (1994). “On Being Systematically Connectionist,” *Mind and Language*, 9(3): 288–302
- Papineau, D. (1987). *Reality and Representation*, Oxford, UK: Basil Blackwell.
- Perry, John and David Israel (1991). “Fodor and Psychological Explanations” in *Meaning in Mind: Fodor and his Critics*, B. Loewer and G. Rey (eds.), Oxford, UK: Basil Blackwell, 1991.
- Phillips, S. (2002). “Does Classicism Explain Universality?” *Minds and Machines* 12(3): 423–434.
- Piccinini, G. (2008). “Computers,” *Pacific Philosophical Quarterly*, 89:32 –73.
- Pinker, S., and A. Prince (1988). “On language and connectionism: Analysis of a parallel distributed processing model of language acquisition,” *Cognition* (special issue on Connections and Symbols) 28: 73–193.
- Plate, Tony A. (1998). “Structured operations with distributed vector representations” in Keith Holyoak, Dedre Gentner, and Boicho Kokinov, *Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences*. NBU Series in Cognitive Science. New Bugarian University, Sofia.
- Pollack, J.B. (1990). “Recursive Distributed Representations,” *Artificial Intelligence*, Vol.46, Nos.1–2, (Special Issue on Connectionist Symbol Processing).
- Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA, MIT Press.
- Putnam, Hilary (1988), *Representation and Reality*, Cambridge, Massachusetts: MIT Press.

- Pylyshyn, Z.W. (1978). "Imagery and Artificial Intelligence" in *Perception and Cognition*. W. Savage (ed.), University of Minnesota Press. (Reprinted in *Readings in the Philosophy of Psychology*, N. Block (ed.), Cambridge, Massachusetts: MIT Press, 1980.)
- Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*, Cambridge, Massachusetts: MIT Press.
- Ramsey, F.P. (1931). "General Propositions and Causality" in *The Foundations of Mathematics*, New York: Harcourt Brace, pp. 237–55.
- Ramsey, W., S. Stich and J. Garon (1991). "Connectionism, Eliminativism and the Future of Folk Psychology," in *Philosophy and Connectionist Theory*, W. Ramsey, D. Rumelhart and Stephen Stich (eds.), Hillsdale, NJ: Lawrence Erlbaum.
- Rescorla, M. (2009a). "Cognitive maps and the language of thought," *The British Journal for the Philosophy of Science*, 60 (2): 377–407.
- –––. (2009b). "Predication and cartographic representation," *Synthese,* 169:175–200.
- Rey, Georges (1981). "What are Mental Images?" in *Readings in the Philosophy of Psychology*, N. Block (ed.), Vol. 2, Cambridge, Massachusetts: Harvard University Press, 1981.
- –––. (1991). "An Explanatory Budget for Connectionism and Eliminativism" in *Connectionism and the Philosophy of Mind*, Terence Horgan and John Tienson (eds.), Studies in Cognitive Systems (Volume 9), Dordrecht: Kluwer Academic Publishers.
- –––. (1992). "Sensational Sentences Switched", *Philosophical Studies* 67: 73–103.
- –––. (1993). "Sensational Sentences" in Consciousness, M. Davies and G. Humphrey (eds.), Oxford, UK: Basil Blackwell, pp. 240–57.
- –––. (1995). "A Not 'Merely Empirical' Argument for a Language of Thought," in *Philosophical Perspectives* 9, J. Tomberlin (ed.), pp. 201–222.
- –––. (1997). *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Oxford, UK: Basil Blackwell.
- Rosenthal, D.M. (1997). "A Theory of Consciousness" in *The Nature of Consciousness: Philosophical Debates*, edited by N. Block, O. Flanagan and G. Güzeldere, Cambridge, Massachusetts: MIT Press.
- Roth, M. (2005). "Program Execution in Connectionist Networks," *Mind & Language*, 20(4): 448–467.
- Rumelhart, D.E. and J.L. McClelland (1986). "PDP Models and General Issues in Cognitive Science," in *Parallel Distributed Processing*, Vol.1, D.E. Rumelhart, J.L. McClelland, and the PDP Research Group, Cambridge, Massachusetts: MIT Press, 1986.
- Rumelhart, D.E., J.L. McClelland, and the PDP Research Group (1986). *Parallel Distributed Processing*, (Vols. 1&2), Cambridge, Massachusetts: MIT Press.
- Rupert, R. D. (1999). "On the Relationship between Naturalistic Semantics and

Individuation Criteria for Terms in a Language of Thought," *Synthese*, 117: 95–131.

- –––. (2008). "Frege's puzzle and Frege cases: Defending a quasi-syntactic solution," *Cognitive Systems Research*, 9: 76–91.
- Sanjeevi, S. and P. Bhattacharyya (2010). "Connectionist predicate logic model with parallel execution of rule chain" in *Proceedings of the International Conference and Workshop on Emerging Trends in Technology* (ICWET 2010) TCET, Mumbai, India (2010).
- Schiffer, Stephen (1981). "Truth and the Theory of Content" in *Meaning and Understanding,* H. Parret and J. Bouveresse (eds.), Berlin: Walter de Gruyter, 1981.
- Searle, John R. (1980). "Minds, Brains, and Programs" *Behavioral and Brain Sciences* III, 3: 417–24.
- –––. (1984). *Minds, Brains and Science*, Cambridge, Massachusetts: Harvard University Press.
- –––. (1990). "Is the Brain a Digital Computer?", *Proceedings and Addresses of the APA,* Vol. 64, No. 3, November 1990.
- –––. (1992). *The Rediscovery of Mind*, Cambridge, Massachusetts: MIT Press.
- Sehon, S. (1998). "Connectionism and the Causal Theory of Action Explanation." *Philosophical Psychology* 11(4): 511–532.
- Shastri, L. (2006). "Comparing the neural blackboard and the temporal synchrony-based SHRUTI architecture," *Behavioral and Brain Science*, 29: 84–86.
- Shastri, L. and A. Ajjanagadde (1993). "From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony," *Behavioral and Brain Sciences*, Vol. 16, pp. 417–94
- Shepard, R. and Cooper, L. (1982). *Mental Images and their Transformations*. Cambridge, Massachusetts: MIT Press.
- Smolensky, Paul (1988). "On the Proper Treatment of Connectionism," *Behavioral and Brain Sciences* 11: 1–23.
- –––. (1990a). "Connectionism, Constituency, and the Language of Thought" in *Meaning in Mind: Fodor and His Critics*, B. Loewer and G. Rey (eds.), : Oxford, UK: Basil Blackwell, 1991.
- –––. (1990b). "Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems," *Artificial Intelligence*, Vol. 46, Nos. 1–2, (Special Issue on Connectionist Symbol Processing), November 1990.
- –––. (1995). "Constituent Structure and Explanation in an Integrated Connectionist/Symbolic Cognitive Architecture" in *Connectionism: Debates on Psychological Explanation*, C. Macdonald and G. Macdonald (eds.), Oxford, UK: Basil Blackwell, 1995.
- Schneider, S. (2009). "The Nature of Symbols in the Language of Thought," *Mind and Language*, 24(5): 523–553.
- Stalnaker, Robert C. (1984). *Inquiry*, Cambridge, Massachusetts: MIT Press.

- Sterelny, K. (1986). “The Imagery Debate”, *Philosophy of Science* 53: 560–83. (Reprinted in *Mind and Cognition,* W. Lycan (ed.), Oxford, UK: Basil Blackwell, 1990.)
- –––. (1990). *The Representational Theory of Mind*, Cambridge, Massachusetts: MIT Press.
- Stich, Stephen (1983). *From Folk Psychology to Cognitive Science: The Case against Belief*, Cambridge, Massachusetts: MIT Press.
- Tarski, Alfred (1956). “The Concept of truth in Formalized Languages” in *Logic, Semantics and Metamathematics*, J.Woodger (trans.), Oxford, UK: Oxford University Press.
- Touretzky, D.S. (1990). “BoltzCONS: Dynamic Symbol Structures in a Connectionist Network,” *Artificial Intelligence*, Vol. 46, Nos. 1–2, (Special Issue on Connectionist Symbol Processing).
- Tye, M. (1984). “The Debate about Mental Imagery”, *Journal of Philosophy* 81: 678–91.
- –––. (1991). *The Imagery Debate*, Cambridge, Massachusetts: MIT Press.
- Van Der Velde, F. and Marc De Kamps (2006). “Neural blackboard architectures of combinatorial structures in cognition,” *Behavioral and Brain Sciences*, Vol. 29 (01), pp. 37–70.
- van Gelder, Timothy (1989). “Compositionality and the Explanation of Cognitive Processes”, *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society*, Ann Arbor, Michigan, pp. 34–41.
- –––. (1990). “Compositionality: A Connectionist Variation on a Classical Theme,” *Cognitive Science*, Vol. 14.
- –––. (1991). “Classical Questions, Radical Answers: Connectionism and the Structure of Mental Representations” in *Connectionism and the Philosophy of Mind*, Terence Horgan and John Tienson (eds.), Studies in Cognitive Systems (Volume 9), Dordrecht: Kluwer Academic Publishers.
- Vinueza, A. (2000). “Sensations and the Language of Thought.” *Philosophical Psychology* 13(3): 373–392.
- Wermter, S. and Ron Sun (eds.) (2000). *Hybrid Neural Systems*, Heidelberg: Springer.

Other Internet Resources

- Bibliography on the language of thought hypothesis, in *MindPapers*, edited by D. Chalmers and D. Bourget.
- Bibliography on the philosophy of artificial intelligence, in *MindPapers*, edited by D. Chalmers and D. Bourget.

Related Entries

artificial intelligence | belief | Church-Turing Thesis | cognitive science | computation: in physical systems | concepts | connectionism | consciousness: representational theories of | folk psychology: as a theory | functionalism | intentionality | mental content: causal theories of | mental imagery | mental representation | mind: computational theory of | naturalism | physicalism | propositional attitude reports | qualia | reasoning: automated | Turing, Alan | Turing machines