Jerry Fodor and the Representational Theory of Mind

Final draft in *Philosophy of Mind: The Key Thinkers*, ed. Andrew Bailey. Bloomsbury (2014).

Matthew Katz
Central Michigan University

Introduction

Among the most important developments in philosophy of mind in the twentieth century was the introduction of the idea that the mind is a computer. For this idea offers a physicalist theory of the mind that neither identifies mental states with patterns of behavior, nor with types of brain states. Perhaps more than any other single philosopher, Jerry Fodor has advanced a version of this idea. More specifically, Fodor has described and defended a collection of related hypotheses he refers to as the *Representational Theory of Mind* (RTM). The most famous of these is the *Language of Thought Hypothesis*, which is the idea that thinking takes place within a mental language, though RTM also consists in the idea that the *propositional attitudes* are relations between subjects and mental representations, and in the idea that thinking is a computational process. Fodor is also well known for offering an account of the overall functional architecture of the mind, according to which the mind is partly composed of modules that perform individualized tasks. Ironically, Fodor argues that his modularity thesis suggests limitations on the extent to which cognition can be explained in terms of computation.

Folk Psychology, Propositional Attitudes, and Functionalism

One way in which questions about the relation between mind and body get expressed is in terms of *folk psychology* and the *propositional attitudes*. Folk psychology, roughly speaking, is the large collection of commonsense laws about the relations between the content of one's mind and one's behavior, which we often use in the explanation and prediction of behavior. For example, we might explain Julie's carrying an umbrella with her to work by citing the presence of storm clouds, Julie's subsequent belief that it might rain, her belief that an umbrella will keep her dry in the rain, and her desire to stay dry. This explanation relies on, among other things, the commonsense general law that normally functioning people see storm clouds when such clouds are present, and form the belief that it might rain.

Such explanations and predictions can make reference to many different sorts of mental phenomena: beliefs, desires, hopes, fears, worries, intuitions, and so on. These mental phenomena are often called *propositional attitudes*, because we express the content of a person's mind in terms of their taking an attitude toward some proposition. For instance, "Julie hopes that it will not rain" claims that Julie holds the attitude of hoping toward the proposition 'It will not rain".

One major component of Fodor's work has been to defend the reality of the propositional attitudes (and thus the truth, for the most part, of folk psychology) against claims that they do not exist (and thus that folk psychology is a largely false theory). His argument relies on the fact that folk psychological explanation and prediction is ubiquitous, that it has been so for millennia, and that it often proves very reliable. Consider one of his examples:

> Someone I don't know phones me at my office in New York from… Arizona. 'Would you like to lecture here next Tuesday?' are the words he utters. 'Yes, thank you. I'll be at your airport on the 3p.m. flight' are the words that I reply. That's *all* that happens, but it's more than enough; the rest of the burden of predicting behavior… is routinely taken up by theory. And the theory works so well that several days later (or weeks later, or months later, or years later…) and several thousand miles away, there I am at the airport, and there he is to meet me. (1987, 3).

The theory Fodor mentions here includes general laws such as "if someone utters the words 'I'll be at your airport on the 3p.m. flight' then all things being equal s/he intends to be at your airport on the 3p.m. flight" (1987, 3) and "all else being equal, people do what they intend to do", and so on. Since our predictions based on the theory work so well so often, Fodor contends, that theory must be true.

One worry some philosophers have had is that folk psychology actually fails to provide any explanation at all of much of our mental lives or behavior. For instance, Paul Churchland asks us to,

> consider the nature and dynamics of mental illness, the faculty of creative imagination, or the ground of intelligence differences between individuals. Consider our utter ignorance of the nature and psychological functions of sleep… Reflect on the common ability to catch an outfield fly ball on the run, or hit a moving car with a snowball… On these and many other mental phenomena, [folk psychology] sheds negligible light. (1981, 73).

Because in his view folk psychology is so woefully inadequate at capturing many of the details of our mental lives, Churchland argues that it must for the most part be false, that it "must inevitably fail to capture what is going on [in our mental lives], though it may reflect just enough superficial structure to sustain an alchemylike tradition among folk who lack any better theory (1981, 85). He maintains, therefore, that folk psychology and its vocabulary will eventually be replaced by neuroscience and its vocabulary.

Many agree though, that the promise of a psychology that eschews traditional psychological terms in favor of neuroscientific ones is still just that, a promise, which may never come to fruition. Fodor, for one, denies that we will ever make good on that promise, insisting that not only is folk psychology incredibly predictive of human behavior, but that so far nothing else is. He writes that

> the predictive adequacy [of folk psychology] is beyond rational dispute… If you want to know where my physical body will be next Thursday, mechanics—our best science of middle-sized objects after all, and reputed to be pretty good in its field—*is no use to you at all*. Far the best way to find out (usually, in practice, the *only* way to find out) is: *ask me!* (1987, 6).

For a philosopher inclined toward a physicalist account of the mind though, realism about the propositional attitudes demands an account of their existence in physical terms. Some have argued that when we refer to mental states like propositional attitudes we are actually referring to patterns of behavior or to dispositions to behave in particular ways (see chapter 4). Others have argued that mental states are identical to types of brain states (see chapter 5). In contrast, Fodor has argued that propositional attitudes are relations between subjects and mental representations. According to this view, for example, Julie believes that it will rain just in case Julie has tokened a mental representation that means "it will rain" and Julie bears the relation of belief to that representation. If Julie believes that it will snow, she bears the same relation to a different mental representation (with a different meaning). If Julie hopes that it will rain, she bears a different relation to the same mental representation.

On this account, the difference between propositional attitudes is a difference in relation one holds toward a given representation. For example, if Julie believes it will snow, she is perhaps likely to wear warm shoes and a warm coat if she goes outside. She is perhaps also likely to affirm, if asked, that it will snow, and she is likely to form other related beliefs such as that the roads may become slippery, that the scenery will be pretty later, and so on. If she holds some other attitude toward the proposition that it will snow, she is perhaps likely to behave in different ways and to form different subsequent thoughts. That is, on Fodor's view propositional attitudes are characterized by the kinds of stimuli that cause them and the kinds of effects they cause, where those stimuli and effects include other internal (mental) states. In short, the account is functionalist (see chapters 6 and 7), and indeed Fodor's early work was devoted to describing and defending a functionalist theory of mind, as a middle ground between dualism and behaviorism. (see, e.g., Fodor 1968).

However, if propositional attitudes are defined by what they are caused by and by what they cause, then one needs an account of how they enter causal relationships with other attitudes, stimuli, and behavior. For example, if Julie's belief that is will snow is a relation between Julie and a mental representation that means "it will snow", and if this relation consists in part in its causing Julie to also believe that the landscape will be pretty later, which itself is a relation between Julie and a mental representation that means "the landscape will be pretty later", then there needs to be an account of the processes that take as input the first mental representation and create as output the second. This requirement leads to the idea that thinking is computation, which in turn leads to the Language of Thought Hypothesis.

Computation and the Language of Thought

In order to understand the idea that thinking is computation, it is useful to begin with a description of Turing machines, the idealized computational devices first described by the British mathematician Alan Turing. Indeed, Turing was the first to argue that thinking is computation (see Turing 1950) and as Fodor notes, his own philosophical project is in large measure an attempt to fit Turing's idea together with the idea that folk psychology is largely true (1994, 1-2).

Turing machines are composed of a string of tape (in theory, infinite in both directions) that is divided into segments, each of which may have a symbol written in it, and a "write-read" head, which moves along the tape, reading, erasing, and writing new symbols in the segments. Strings of symbols on the tape may be interpreted in various ways, for example as numbers or words or logical formulae, and the write-read head follows a set of precise directions that tells it when to erase a symbol, when to leave a symbol alone, when to write a symbol, and when to stop. Depending on what directions are given to the write-read head, the machine can solve mathematical problems, perform logical derivations, and so on.

The directions Turing machines follow refer to the syntax (roughly, the structure) of the strings of symbols, and have no understanding or knowledge of semantic features (roughly, the meaning) of those symbols. For example, a machine that doubles integers might represent those integers with strings of the symbol '*' and have directions that instruct it how to erase each * and replace it with two *s, thus doubling the number represented. Still, even though the machine has no knowledge whatsoever that the strings mean *n* and *2n,* respectively, the machine will always give the correct answer of *2n*, when asked to double *n* (provided of course that the directions are correctly written and the machine does not malfunction in some way).

In general, even though Turing machines only pay attention to the structure of the symbols they process, they are able to respect various constraints on the meaning of those symbols such as reliably giving correct answers to mathematical questions. Moreover, because they only pay attention to syntax, they can be implemented as actual physical machines.[1] Indeed, Turing machines are critical progenitors of today's modern digital computer. In today's computing terminology, the tape is the Turing machine's 'memory', the symbols written on the tape are 'data' stored in that memory, the write-read head is its 'central processing unit', and the directions it follows constitute its 'program'.

More importantly for present purposes, the ability of a Turing machine to compute the correct answer to a question, while nevertheless not understanding that it is answering a question, provides a model of how propositional attitudes can cause and be caused by other propositional

[1] Strictly speaking this is not accurate. In theory, a Turing Machine's tape extends infinitely in both directions, and so cannot be built. Nevertheless, the Turing Machine is the basis for the idea of a mechanical device that processes representations syntactically.

attitudes. Specifically, mental representations are akin to the strings of symbols on a Turing machines's tape, and the specific attitude a subject takes toward that representation is akin to the set of directions that the Turing machine follows in performing its computations. In other words, the idea that thinking is computation is the idea that propositional attitudes are sets of instructions that determine what mental representations to token, given other previously tokened representations (and that determine what behaviors should be performed, etc).

Fodor explains that,

> I assume that psychological laws are typically implemented by computational processes… Computational processes are defined over syntactically structured objects… There is a well-known and, in my opinion, completely convincing argument for viewing the implementation of psychological laws in this way: It is characteristic of the mental processes they govern that they tend to preserve semantic properties like truth. Roughly, if you start out with a true thought, and proceed to do some thinking, it is very often the case that the thoughts that the thinking leads you to will also be true. This is, in my view, the most important fact we know about minds… Well as Turing famously pointed out, if you have a device whose operations are transformations of symbols, and whose state changes are driven by the syntactic properties of the symbols that it transforms, it is possible to arrange things so that, in a pretty striking variety of cases, the device reliably transforms true input symbols into output symbols that are also true. I don't know of any other remotely serious proposal for a mechanism that would explain how the processes that implement psychological laws could reliably preserve truth… So I assume that Turing was right: the mind is a computer of some sort or other. (1994, 7-9).

In fact, an appeal to a lack of other possibilities is not Fodor's only argument in favor of a computational view of mental processes. In his landmark book *The Language of Thought* (1975) he argued that (at the time) extant accounts of a variety of psychological processes implicitly assumed that thinking is computational. He argued for instance that according to theories of decision making, subjects settle on a particular course of action by creating a preference ordering over possible consequences of available actions, and computing the likelihood of those consequences (1975, 28-31).

An important part of the idea that thinking is computation is that it implies the existence of a system of mental representation, and therefore fits well with the account of propositional attitudes as relations between subjects and mental representations, which obviously also implies the existence of mental representations. Consider again Turing machines (and modern digital computers): the very idea of them depends on the idea that there are symbols in at least some of the squares on the tape (i.e., that there are bits of data in memory). Otherwise, there is nothing over which to perform computations. Put another way, the program the machine follows is a set of directions about when to write and erase symbols. So without symbols there can be no program. Or as Fodor writes, "according to [theories of decision making], deciding is a computational process; the act the agent performs is the consequence of computations defined over representations of possible actions. No representations, no computations" (1975, 31).

According to Fodor though, the idea that thinking is computation implies not only the existence of a system of mental representation, but also that that system has certain features. In particular, he argues that the system must be language-like in structure. This then, is the famous Language of Thought Hypothesis (LOTH). The idea that mental representations have a language-like structure amounts to the idea that they are composed from a finite store of atomic representations (those that have no meaningful parts, as words in English), which may be combined to form compound representations (as sentences are composed of words). The meaning of a compound representation, moreover, is a function of the meaning and arrangement of its component parts (as the meaning of a sentence is determined by the meaning and arrangement of the words in it).[2]

[2] This description of what linguistic representation amounts to is present in Fodor (1975) and becomes an explicit and central part of the argumentation in Fodor and Pylyshyn (1988).

It is no accident, therefore, that LOTH is sometimes described as the view that thoughts are sentences in the head. This can be a useful way of understanding the theory, but it needs to be clarified as well. To begin, it is important to understand that sentences can be written in many ways in many media. They are written in pen on paper, they are etched in stone, encoded in the dots and dashes of Morse code, and so on. LOTH then, is simply the idea that some of the brain's activity can in principle be analyzable as the encoding and processing of representations that have a linguistic structure.

Also, it is also important to recognize that it is no part of LOTH that these representations are consciously available to thinking subjects. The hypothesis concerns the form of representation employed by the brain, "beneath and behind" what subjects are aware of when they are thinking. This becomes quite clear when considering one of Fodor's early arguments for LOTH: that explaining how young children acquire a first spoken language requires positing a language of thought.

Fodor argued that language acquisition should be seen as a kind of hypothesis formation and confirmation. The idea was that infants and young children create hypotheses about the extension of the words they hear, and test those hypotheses by using those same words to refer to objects around them. These hypotheses are confirmed by affirmation (given by nearby competent speakers) or disconfirmed by correction. Fodor argued, however, that these hypotheses themselves must be cast within an internal representational system with linguistic structure (i.e., a language of thought) and thus, that we cannot explain how human beings acquire natural languages unless we posit a language of thought. And as one would expect, Fodor did not conceive of infants and toddlers as being consciously aware of engaging in hypothesis confirmation (1975, 58-64).

Another reason Fodor gives for thinking that the brain employs a language-like system of representation is that it allows for an infinity of unique representations, a property known as *productivity*. He explains that,

> the essential point is the organism's ability to deal with *novel* situations. Thus, we infer the productivity of natural languages from the speaker/hearer's ability to produce/understand sentences on which he was not specifically trained. Precisely the same argument infers the productivity of the internal representational system from the agent's ability to calculate the behavioral options appropriate to a kind of situation he has never before encountered (1975, 31-2).

But the ability to produce an infinity of unique representations from finite means, argue Fodor and Pylyshyn (1988), demands that arbitrarily complex representations be constructed from a finite store of atomic representations, in which the meaning of any of those complex representations is a function of the meaning and arrangement of the component parts. But that just is a system with linguistic structure.

Similarly, Fodor and Pylyshyn (1988) argued for LOTH from the premise that thought is *systematic*.[3] The idea that thought is systematic is the idea that certain thoughts are related to certain other thoughts in such a way that any thinker who understands the one will also be able to understand the other. For instance, Fodor and Pylyshyn claim that anyone who can entertain the thought *Mary loves John* can also entertain the thought *John loves Mary*. They argue that the best explanation for this is that each is a compound representation, that the compounds share all the same parts and differ only in the arrangement of those parts, and that understanding either implies understanding the component parts. Therefore, if a subject is capable of understanding the one, she will also be capable of understanding the other, since doing so requires all the same abilities as understanding the first. But if this is what explains why the thoughts are

[3] They added an argument from *inferential coherence* as well. Thought is inferentially coherent just in case given that a subject is capable of drawing some instances of a kind of logical inference, she is able to draw any instance of that kind of logical inference.

systematicallly related, then LOTH must be true, for the idea that the representations are composed of component parts and have meanings that are dependent on those parts and their arrangements, again, just is the idea that they have a linguistic structure.

Significance and Objections

The significance of RTM is often best seen in relation to Descartes' famous claim that a machine could not use reason. He writes that, "it is not conceivable that… a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do" and that it would be impossible for a machine to "act in all the contingencies of life in the way in which our reason makes us act" (1637/1988 44-5). Descartes took this line of argument to show that the mind is an immaterial substance that, while joined to the body during life, nevertheless has a "nature entirely independent of the body" (1637/1988, 46).

But Turing's work and the subsequent development of the digital computer suggest that Descartes may be exactly wrong here: by following directions that tell a machine how to manipulate symbols based on the structuring of those symbols, the machine can indeed respond in incredibly complex ways, ways that respect truth and other semantic constraints. If one applies this idea to the human brain, the result is the hypothesis is that thinking is computation, and that explaining reason may need no appeal to immaterial substances after all.

The hypothesis that thinking is computation though, implies a system of mental representation, and according to Fodor, that that system has a linguistic structure. Moreover, the idea that thinking is computing over mental representations allows for the claim that mental states such as propositional attitudes are akin to the programs that ultimately determine the behavior of a computer. That is, the attitudes determine what stimuli and representations will cause a given representation, and what other representation and behavior that representation will in turn cause. In short, though dualism is rejected, so too is any identification of mental states with behavior or with types of brain states.

Still, objections to the above constellation of views and in particular to Fodor's presentation of them, are legion. It was noted above that some philosophers have argued that folk psychology is a false theory. Moreover, as a version of functionalism, RTM is susceptible to a number of objections facing such theories (see chapters 6 and 7). Many philosophers have argued that any version of computationalism is bound to fail, or at least, that it faces technical difficulties so challenging they may in fact be insurmountable (see for example Searle 1980 and Dennett 1984). LOTH itself has been widely controversial. Some philosophers and psychologists have argued that rather than having a linguistic structure, much mental representation has an imagistic format (e.g., Kosslyn 1980). Others have suggested that some mental representation is map-like (e.g., Braddon-Mitchell and Jackson 1996). Still others have wondered whether thought really does have the properties of productivity and systematicity, as Fodor and Pylyshyn claim it does (e.g., Johnson 2004).

Many philosophers initially rejected LOTH because it was seen to be wedded to a particularly strong version of the idea that some concepts are possessed innately. In particular, Fodor's original arguments for LOTH were coupled with an argument that all atomic concepts (i.e., "words" in the language of thought) are innate. The argument for the latter claim was that if one were to learn an atomic concept, it would have to be by way of hypothesis confirmation (akin to that which explains language acquisition, as discussed above), but that such a process would presuppose knowledge of the concept to be learned and is thus impossible. Fodor argued for instance that if one is to learn the concept RED, then she must be able to form hypotheses about which things are red and which things are not. But such hypotheses must be couched in a system of representation that has resources to express propositions about which things are red and which are not. That is, the system must already possess the concept RED. So one cannot learn the concept RED unless one already possesses it. So learning RED is impossible, and if anyone has the concept they must have possessed it innately.

However, a list of atomic concepts will include many that seem absurd to suppose are possessed innately—CARBURETOR has been a favorite in the literature—and many philosophers took that as argument that Fodor's view is false, or indeed, absurd (e.g., Putnam 1988, Churchland, 1986). Fodor himself admits that "This conclusion, according the consensus, was downright loopy" (2008, 129), and though he has not abandoned the view he has refined it. While the reasoning in *The Language of Thought* assumed that if a concept is not learned, then it must be innate, Fodor has since distinguished between acquiring a concept and learning one, claiming that learning a concept is just one way among many others of acquiring a concept. He writes that,

> *There are all sort of of mind/world interactions that can alter a conceptual repertoire. Concept learning* (if there is such a thing) *is one, but it's certainly not the only one.* Other candidates conceivably include: sensory experience, motor feedback, diet, instruction, first-language acquisition, being hit on the head by a brick, contracting senile dementia, arriving at puberty, moving to California, learning physics, learning Sanskrit, and so forth indefinitely. (2008, 131-2).

He concludes that while the argument that concepts cannot be learned still stands, it does not follow that all concepts are innate. Rather, the question becomes how organisms acquire concepts, since they clearly do so but neither learn them nor possess them innately. He writes,

> The central issue isn't *which concepts are learned*, since… none of them are. Nor, however, is it *which concepts are innate*, if an innate concept is one the acquisition of which is independent of experience. Quite likely, there are none of those either. Rather, the problem is to explain how a creature's innate endowment (whether it is described in neurological or in intentional terms) contributes to the acquisition of its conceptual repertoire; that is, how innate endowments contribute to the processes that start with experience and end in concept possession. (2008, 145).

One very well-known objection to LOTH stems from the advance in the 1980s of *connectionist networks*. When Fodor wrote *The Language of Thought*, the Turing model of computing was the only available model, and as the model demands linguistically-structured representations, so Fodor argued that viewing the mind as a computer demands positing a language of thought. But in the 1980s cognitive scientists began to have serious successes modeling various cognitive processes using connectionist networks, computational systems that do not employ linguistically-structured representations (see chapter 11, and Churchland 1995). Moreover, connectionist networks appear to resemble the brain both in gross architecture and functioning more so than Turing machines and digital computers. They are composed of groups of simple processing units and connections among the units, they lack both dedicated memory and a central processing unit, and activity passes through them in massively parallel fashion, as in the brain, and as opposed to the step-by-step serial processing found in digital models. So some authors have argued that if the brain is a computer, it is one much more akin to such networks than to Turing machines and digital computers.

Much of this debate concerns the possibility that connectionist networks do not offer an alternative to LOTH, but rather suggest a way in which the language of thought might be implemented in the brain. Fodor and Pylyshyn (1988) and Fodor and McLaughlin (1990), for instance, argue that because network architectures lack linguistically-structured representations, they cannot account for the productivity and systematicity of thought. Therefore, they argue, they cannot offer a genuine alternative to LOTH. Still others like Smolensky (1988) have tried to show that connectionist networks can indeed explain those phenomena.

One intriguing aspect to debates surrounding the idea that thought is computation is that Fodor himself has argued that it should only be taken to explain the functioning of certain parts of the mind. This argument rests on a view of the overall architecture of the mind that Fodor has defended.

Modularity and the Limits of Computation

In *The Modularity of Mind* (1983), Fodor argues that a general functional taxonomy of psychological systems would include three kinds of system. This taxonomy would include systems that take in raw data from the environment, called *transducers*, systems that interpret that data, called *input systems*, and systems for everything else, including higher cognitive functions such as belief formation, called *central systems*. Fodor argued that if this taxonomy is correct, it shows that there are significant limits on the explanatory power of the computational aspect of RTM. He writes,

> Over the years, I've written a number of books in praise of the Computational Theory of Mind… It is, in my view, by far the best theory of cognition that we've got… There is… every reason to suppose that the Computational Theory is part of the truth about cognition… But it hadn't occurred to me that anyone could think that it's a very *large* part of the truth; still less that it's within miles of being the whole story about how the mind works (2000, 1).

Fodor's notion of a transducer is that of a system whose output is "most naturally interpreted as specifying the distribution of stimulations at the 'surfaces' (as it were) of the organism" (42) in a format that may then be used by the input systems. For example, it might be the job of a transducer to present patterns of retinal stimulation to the input systems in a format the latter systems employ. The input systems then, "deliver representations that are most naturally interpreted as characterizing the arrangement of *things in the world*" (42). Those representations may then be used by the central systems for all manner of higher cognitive functions.

The central claim of *Modularity* is that the input systems, and not the central systems, should be thought of as modules, performing specific tasks that can be more-or-less separated from the rest of the system (roughly like the way a compact disc player might be removed from a stereo system, leaving the rest of the system functioning normally). The argument rests on a list of features that Fodor believes is characteristic of modules, on reasons to believe input systems possess those features, and on reasons to believe that central systems do not. For present purposes the most important of these features are *domain specificity* and *informational encapsulation*.

To say that input systems are domain specific is to say that "the range of distal properties they can project… hypotheses about" is quite narrow (47). For instance, "Candidates might include, in the case of vision, mechanisms for color perception, for the analysis of shape, and for the analysis of three-dimensional spatial relations" (47). To say that input systems are informationally encapsulated is to say that their operations are not affected by feedback of "information that is specified only at relatively high levels of representation" (64). In other words, while the products of input systems are taken as input for computations performed by central systems, the reverse is not true: the products of central systems are not used in computations performed by input systems. Fodor takes perceptual illusions to be evidence of this claim. For example,

> The very same subject who can tell you that the Muller-Lyre arrows are identical in length, who indeed has seen them measured, still finds one looking longer than the other.

> In such cases it is hard to see an alternative to the view that at least *some* of the background information at the subject's disposal is inaccessible to at least some of his perceptual mechanisms. (66)

Although Fodor describes the function of input systems as characterizing things in the world, he also thinks that their outputs are *shallow*, in the sense that they only employ *basic* categories in doing so, where by "basic" he suggests categories that are neither very abstract nor very particular (94). For example, he suggests that input systems might group objects into the category *dog*, but not *poodle* or *thing*. Of course, cognition involves a great deal more than grouping objects into basic categories. Among so many other things, it also involves both more specific and more abstract categorization, belief and hypothesis formation, problem solving, and so on. According to Fodor then, the central systems are responsible for all these other functions that neither transducers nor input systems perform.

But if that is true, Fodor argues, then central systems must, in stark contrast to input systems, be both (relatively) domain general and informationally un-encapsulated. To begin with, since input systems are domain specific, for example forming representations of the color of objects in the environment or of the structure of a rhythm, yet we form beliefs about the relationship between our visual and auditory experiences, the systems responsible for forming those beliefs will have to cross visual and auditory domains. That is, they will have to be relatively domain general as contrasted with input systems. Moreover, Fodor argues that central systems must be un-encapsulated. The line of reasoning here is that we often reason analogically—borrowing information from one domain to use in another—and such reasoning by definition cannot be encapsulated. Here Fodor looks to the history of scientific discovery as being rich with examples:

> what's known about the flow of water gets borrowed to model the flow of electricity; what's known about the structure of the solar system gets borrowed to model the structure of the atom; what's known about the behavior of the market gets borrowed to model the process of natural selection, which in turn gets borrowed to model the shaping of operant responses… The point about all this is that 'analogical reasoning' would seem to be… a process which depends precisely upon the transfer of information among cognitive domains previously assumed to be mutually irrelevant. By definition, encapsulated systems do not reason analogically. (1983, 107)

Fodor's account of the mental architecture is important in its own right, has been widely influential, and other accounts of have been offered in its wake. Some philosophers have argued that there are indeed the modules Fodor says there are, plus other specific ones. Currie and Sterelny (2000) for example, suggest a module for social cognition. Still others have argued that the mind is massively modular, being composed of perhaps hundreds of thousands of domain specific modules, each one designed by natural selection to perform a function that would have aided our evolutionary ancestors in some way (e.g., Cosmides and Tooby 1994).

Fodor takes his account of modularity to indicate that an appeal to computation is limited in its ability to explain cognition, specifically because the central systems are domain general and un-encapsulated. He argues that the upshot of domain general un-encapsulated central systems is that "there seems to be no way to delimit the sorts of informational resources which may affect, or be affected by central processes… We can't, that is to say, plausibly view the fixation of belief as effected by computations over bounded, local information structures" (112). In other words, forming beliefs appears to involve computations that are sensitive to everything the organism knows. Unfortunately, according to Fodor, we have no understanding of and no way to characterize such "global" sorts of computations. As noted above, his notion of computation is that of processes that are defined over the syntactic structure of representations, and syntactic structure is a local—not global—property. That is, knowing the syntactic structure of a representation does not require knowing anything about other representations. Thus, since Fodor's idea of computation does not fit with his view about how the central systems operate, he concludes that his view of computation cannot be employed in accounts of central cognitive processing. And because he thinks his view of computation is the best available theory to employ in accounts of cognition in general, he takes this result to be "very bad news for cognitive science" (1983, 128).[4]

Conclusion

Regardless of the position one takes with respect to the various parts of RTM, their collective significance is difficult to overstate. Together they comprise answers to some of the oldest and thorniest philosophical problems. Those answers are that our everyday descriptions of our minds and the minds of others are largely true, and that psychological explanation will inevitably rest on mental terms, but that the mind is entirely physical, yet not to be identified with patterns of behavior or types of brain states. Rather, mental states are functional states—indeed they are computational states—relating subjects with inner mental representations. Those representations, moreover, together form a language.

Still, even if this theory is true, it is merely the sketch of a complete theory of the nature of thought and thinking. The details of the representational structures involved, as well as the details of the processing employed, and a myriad of other aspects of the system all remain to be described. Moreover, if Fodor's account of modularity is correct and his argument concerning central systems is sound, then a computational account may only be possible for those parts of the mind take in information about the environment and create representations that may then be

[4] Of course, these arguments are hotly disputed as well. If the mind is composed entirely of modules, then computationalism may survive. Alternatively, some have argued that the problems Fodor describes here (and similarly in his 2000), do not in fact apply (e.g., Ludwig and Schneider 2008).

employed by higher faculties. What happens after that—how those higher faculties process those representations—may remain entirely mysterious.

References

Braddon-Mitchell, D., and F. Jackson. (1996). *Philosophy of Mind and Cognition*. Oxford, UK: Blackwell.

Churchland, P.M. (1981). "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* v. 78 n. 2: 67-90.

Churchland, P. M. (1995). *The Engine of Reason, the Seat of the Soul*. Cambridge, MA: MIT Press.

Churchland, P.S. (1986). Neurophilosophy. Cambridge, MA: MIT Press.

Cosmides, L. and J. Tooby. (1994). "Origins of Domain Specificity: The evolution of Functional Organization." In *Mapping the Mind: Domain Specificity in Cognition and Culture*, eds. Hirschfeld, L., and S. Gelman. Cambridge, UK: Cambridge University Press. 85-116.

Currie, G. and K. Sterelny. (2000). "How to think about the modularity of mind reading." *Philosophical Quarterly* v. 50 n. 199: 145-60.

Dennett, D. (1984). "Cognitive Wheels: The Frame Problem of AI." In *Minds, Machines, and Evolution: Philosophical Studies*, ed. Hookway, C. Cambridge, UK: Cambridge University Press. 129-52.

Descartes, R. (1637/1988). "Discourse on the Method." In *Descartes, Selected Philosophical Writings*, trans. Cottingham, J., R. Stoothoff, and D. Murdoch. Cambridge, UK: Cambridge University Press.

Fodor, J. (1968). *Psychological Explanation: An Introduction to the Philosophy of Psychology*. New York, NY: Random House.

______. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

______. (1983) *The Modularity of Mind*. Cambridge, MA: MIT Press.

______. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind.* Cambridge, MA: MIT Press.

______. (1994). *The Elm and the Expert: Mentalese and its Semantics*. Cambridge, MA: MIT Press.

______. (2000) *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.

______. (2008). *LOT 2: The Language of Thought Revisited*. Oxford, UK: Oxford University Press.

Fodor, J. and B. McLaughlin. (1990). “Connectionism and the Problem of Systematicity: Why Smolensky’s Solution Doesn’t Work.” *Cognition* v. 35 n. 2: 183-204.

Fodor, J. and Z. Pylyshyn. (1988). “Connectionism and Cognitive Architecture: A Critical Analysis.” *Cognition* v. 28 n. 1: 3-71.

Johnson, K. (2004). “On the Systematicity of Thought and Language.” *Journal of Philosophy* v.101, no. 3: 111-39.

Kosslyn, S. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

Ludwig K, and S. Schneider. (2008). “Fodor’s critique of the classical computational theory of mind.” *Mind and Language* v. 23 n. 1: 123-43.

Putnam, H. (1988). *Representation and reality*. Cambridge, MA: MIT Press.

Searle, J. (1980). “Minds, Brains, and Programs.” *Behavioral and Brain Sciences* v. 3 no. 3: 417-57.

Smolensky, P. (1988). “The Constituent Structure of Mental States.” *Southern Journal of Philosophy* v. 26 n. s1: 137-60.

Turing, A. (1950). “Computing Machinery, and Intelligence.” *Mind* v. 59 n. 236: 433-60.