Regress Arguments against the Language of Thought

Stephen Laurence and Eric Margolis

Abstract. The Language of Thought Hypothesis is often taken to have the fatal flaw that it generates an explanatory regress. The language of thought is invoked to explain certain features of natural language (e.g., that it is learned, understood, and is meaningful), but according to the regress argument, the language of thought itself has these same features and hence no explanatory progress has been made. We argue that such arguments rely on the tacit assumption that the entire motivation for the language of thought consists in explaining the explanandum that allegedly generates the regress. But this tacit assumption is simply false. The Language of Thought Hypothesis is a cogent view and one with considerable explanatory advantages.

The Language of Thought Hypothesis (LOT) is a familiar theory in contemporary philosophy of mind, though it remains highly controversial.[1] One of the most persistent and influential reasons for rejecting the theory is that it involves some kind of infinite regress. Indeed, in many circles the Regress Argument (as we will call it here) is considered decisive. Defenders of LOT, we are told, have failed to appreciate a pivotal turn in analytic philosophy and are repeating mistakes that Wittgenstein and others have warned us about time and again. Simon Blackburn's discussion in *Spreading the Word* is fairly representative. While endorsing the Regress Argument he suggests that it forms a central part of a set of 'considerations which are by now quite familiar in modern philosophy of language' that 'destroy' any 'dog-legged theory'—Blackburn's term for a theory that holds that words are 'reinterpreted into another medium, such as that of Ideas, whose own powers explain the significance words take on' (Blackburn 1984: 40).

We think this is all wrong. But, moreover, we think that even the most ardent supporters of LOT tend to give the Regress Argument too much credit. The argument is fundamentally mistaken. To show this, we will work through three crucial versions of the argument and point to the different ways supporters of LOT have handled them. Then we will offer our own response.

---

[1] LOT claims that much of cognition takes place in an internal system of representation that has language-like structure in the sense that it has a compositional syntax and semantics. It is not part of LOT, as we will be construing it here, that the internal system of representation is innate, species universal, or even distinct from (in the sense of being non-isomorphic to) natural languages. Likewise, LOT does not involve the claim that there is a single system of mental representation that is used in all human cognition, nor does it involve the claim that every aspect of a mental life can be explained by reference to a language of thought.

Let's begin with Jerry Fodor's discussion in what has come to be the classic presentation and defence of LOT (in Fodor 1975). Fodor considers two versions of the Regress Argument, the first of which addresses the fact that natural languages are learned. The argument can be represented as follows.

Regress on Learning

(1) Natural languages are learned.

(2) Supporters of LOT appeal to certain features of a postulated language of thought in order to explain this fact.

(3) But the language of thought must also be learned, so supporters of LOT must now explain how we learn this internal language.

Dilemma:

(4) Either, the learning of this internal language is explained in the same way the learning of natural language is explained, in which case another language will have to be invoked—i.e., a third language—and an infinite regress ensues.

(5) or the learning of this internal language is explained in some other way, in which case this alternative explanation might have been given for natural language, and the introduction of a language of thought could have been avoided.

The bottom line is supposed to be that defenders of LOT either find themselves in an infinite regress or else their commitment to a language of thought is gratuitous. Fodor's notorious response is to deny premiss (3). According to Fodor, the language of thought isn't learned—it's innate—so there isn't a regress.

This response may not be entirely wrong, but we aren't nearly as comfortable with it as Fodor seems to be. First, it relies upon a strong empirical hypothesis—the innateness of the language of thought—which is considered highly dubious amongst people who otherwise are sympathetic to LOT. Clearly, a better response would leave open the question of the extent to which the language of thought is innate. Second, Fodor's response doesn't generalize to other versions of the Regress Argument, versions we will come to shortly. But while different versions of the Regress Argument may require individual responses, their common structure suggests there may well be a common problem with them.

The second version of the Regress Argument that we want to consider, also discussed by Fodor, turns on our ability to understand natural language. As Fodor points out, he might be able to deny that the language of thought is learned, but he can't really deny that 'it is, in a certain sense, understood' (1975: 65). Focusing on understanding, we get the following version of the Regress Argument:

Regress on Understanding

(1) Natural languages are understood.

(2) Supporters of LOT appeal to certain features of a postulated language of thought in order to explain this fact.

(3) But the language of thought must also be understood, so supporters of LOT must now explain how we understand this internal language.

Dilemma:
(4) Either, the understanding of this internal language is explained in the same way the understanding of natural language is explained, in which case another language will have to be invoked—i.e., a third language—and an infinite regress ensues;
(5) or the understanding of this internal language is explained in some other way, in which case this alternative explanation might have been given for natural language, and the introduction of a language of thought could have been avoided.

Fodor's response this time is to opt for the second horn of the dilemma. He claims that the internal language is understood in a different sense than natural language is. His hypothetical interlocutor, however, presses him further:

> 'you admit that there is at least one language whose predicates we understand without the internal representation of truth conditions. ... This saves you from infinite regress, but it suggests that even the regress from the natural language to the inner language is otiose. You argue that we learn 'is a chair' only if we learn that it falls under the truth rule '*y is a chair' is true iff x is G* and then you say that the question of learning a truth rule for *G* doesn't arise. Why not stop a step sooner and save yourself the trouble? Why not say that the question of how we learn 'is a chair' doesn't arise either? Explanation has to stop somewhere'. (1975: 66-7)

His response is that,

> explanation has to stop somewhere but it doesn't have to—and it better not—stop *here*. The question of how we learn 'is a chair' does arise precisely because English is learned. The question of how *G* is learned does not arise precisely because, by hypothesis, the language in which *G* is a formula is innate. (1975: 67)

Notice that Fodor has slipped back to the first version of the Regress Argument, the one having to do with language learning. What he hasn't done is answer the charge that 'the regress from the natural language to the inner Language is otiose' specifically in the case of language understanding, the case at hand.

The third version we want to consider is the one that Blackburn has in mind and one that has been recently criticized by Tim Crane (in Crane 1995).This time, the argument turns on the semantic properties of natural language utterances.

Regress on Meaning
(1) Natural language utterances are meaningful. (There are two ways to read this claim. The first concerns the fact that linguistic expressions have any content at all, while the second concerns the fact that a particular expression, or expression type, has a particular content. For present purposes, this distinction doesn't matter, though the reading Crane and Blackburn have in mind is clearly the second.)
(2) Supporters of LOT appeal to certain features of a postulated language of thought in order to explain this fact.

(3) But expressions in the language of thought are also meaningful, so supporters of LOT must now explain how this is so.

Dilemma:

(4) Either, the semantical properties of this internal language are explained in the same way as the semantical properties of natural language are explained, in which case another language will have to be invoked—i.e., a third language—and an infinite regress ensues;

(5) or the semantical properties of this internal language are explained in some other way, in which case this alternative explanation might have been given for natural language, and the introduction of a language of thought could have been avoided.

Crane's reply is in line with the one Fodor gives to the version based on language understanding. Crane says that representations in the language of thought 'have their meaning in a very different kind of way to the way public language sentences do' (151). This, he says, 'does avoid the objection. But now of course, the question is: how *do* Mentalese sentences get their meaning? '(151). But this *doesn't* answer the question, because, just as with Fodor's response, nothing has been said to meet the charge that 'the regress from the natural language to the inner language is otiose'.

It's important to see that this aspect of the second horn of the dilemma is crucial to the Regress Argument. Without it, the Regress Argument isn't really an argument against LOT at all; it merely points out that we need to give different accounts of how the language of thought is learned, understood, or meaningful than we do for natural language—hardly enough to, in Blackburn's words, 'destroy' LOT. In other words, the force of the Regress Argument comes from the suggestion that the language of thought is unmotivated if it *doesn't* lead to an infinite regress, that we might just as well apply to natural language whatever account works for the language of thought and avoid the detour through the language of thought.

We should say that we do have considerable sympathy with the common reductive strategy embodied in the replies given by Fodor and Crane as a way of providing an adequate account of, e.g., the meaningfulness of linguistic expressions. The best theory of linguistic meaning may well be one that appeals to mental representations with semantic properties.[2] Still, we think this response grants far too much *as a response to the Regress Argument*. It simply wouldn't follow that LOT is gratuitous even if the explanation of how the language of thought is meaningful could be applied directly to the problem of how natural language is meaningful. At best, what would follow is that the particular fact in question—that natural language is meaningful—is by itself insufficient to motivate LOT. In short, the Regress Argument tacitly supposes that what drives the language of thought theorist from the level of natural language to the level of a language of thought are exactly the explananda it addresses, that the language of thought is invoked to account for

---

[2] One reason to believe this is that mental representations are more likely to be governed by relevant causal laws than are utterances in a public language. As Fodor says in a related context, theories of content will often stand 'a much better chance of working for mental representations than … for (e.g.) English words. … [Since] whether an English word gets tokened (e.g., uttered) depends not just on what it means but also upon the motivations, linguistic competences, and communicative intentions of English speakers. Giving voice to an utterance, unlike entertaining a thought, is typically a voluntary act (1987: 99-100).

linguistic meaning, learning, and understanding. Yet while it's true that the language of thought theorist will typically explain these things by appealing to a language of thought, it is certainly not true that the *only* reason she postulates a language of thought is to explain these things. Rather, the situation is that she thinks there are excellent independent grounds for endorsing LOT, empirical arguments, only some of which make contact with issues having to do with natural language. So the proper reply to the Regress Argument can't be just that there are good reasons to go from natural language to the language of thought to solve the problem of linguistic meaning (to take one example), for, even if there weren't, we still wouldn't have an argument against LOT. The reason we wouldn't is because the Regress Argument would first have to be supplemented in a way that rules out all the independent arguments for LOT.

To put the point another way, the dialectical situation is that the language of thought theorist, given her independent reasons for endorsing LOT, has it as an option to reduce certain problems about natural language to corresponding problems about language of thought representations, and then solve them there; she isn't forced to opt for a reductive strategy. She could, for example, apply the same type of solution at both levels by appealing to a use theory of meaning for natural language while simultaneously appealing to a functional role semantics for the language of thought. If she finds that a single solution like this works at both levels, then, for all the Regress Argument shows, she is free to use it at both levels. On the other hand, if she finds that her preferred solution works only at the level of a language of thought, then she can reduce the natural language problem to a language of thought problem, and then solve it there. This doesn't mean that anything goes, of course, and there may well be reasons to prefer a reductive solution, as we already noted. Naturally, the question should be settled by familiar theoretical considerations, including overall theoretical elegance and simplicity. The main point, however, is that the LOT theorist is free to adopt either sort of solution. Either way, she isn't caught in an infinite regress, nor is her commitment to LOT gratuitous, *because her reasons for postulating a language of thought are not the very problems that appear, in parallel, at both level*s.

The standard arguments for the language of thought are mixed in that they support different features of the theory. Some argue only for mentalism, others for intentional realism, still others for internally structured mental representations. What's more, the multitude of arguments on offer are not equally persuasive. Obviously we can't review all of the arguments here. So we'll just note a few of the basic reasons for supposing there is a language of thought. To begin, we should think of human behaviour as mediated by a representational system because this would explain the high degree of freedom that exists between environmental states and behavioural consequences. As Fodor says 'the causal relation between stimulus and response is *typically* mediated by the organism's internal representations of each (1975: 157).[3] Moreover, to explain an organism's ability to reason hypothetically and to deal with novel environmental situations (as such), we need to suppose that the organism has a productive representational system, i.e., one, which under suitable idealizations, is unbounded. And to account for the productivity of this system, we have

---

[3] In addition to offering a defence of LOT against a variety of arguments, Fodor is of course one of the main sources of positive arguments for LOT. Unfortunately, he doesn't use t in his reply to the Regress Argument, perhaps because he thinks he has a perfectly adequate answer already. In any event, many of his readers have taken his actual response to the Regress Argument to be a reductio ad absurdum of LOT, since they've found his 'radical' concept nativism absurd (see, e.g., Patricia Churchland's discussion in sec. 9.6 of her 1986). For this reason it's crucial to see that there is a much more powerful and perfectly general response, which can be given to each version of the Regress Argument, and not just to the argument based on learning.

to suppose that it has a compositional syntax and semantics. This appeal to structure is also required to provide an account of the mechanisms involved in psychological processes. Many psychological processes—lexical insertion, syntactic transformation, phonological encoding, and so on—seem to involve the manipulation of parts of mental representations (see, e.g., Fodor 1987, Levelt 1993). But in order to be able to manipulate parts of representations, representations have to have parts, so they have to be, to some extent, structured representations

In the end, it may happen that these motivations for LOT don't pan out. Maybe some alternative to the language of thought can do all of the same work, or maybe it can be argued that the work doesn't need to be done. But it's important to see that in order for the Regress Argument to have any bite, such arguments would have to be given. The sorts of arguments for LOT that we've briefly mentioned turn out to be crucially involved in the correct evaluation of the Regress Argument. They are independent arguments in the sense that they bring to bear explananda that the Regress Argument ignores, but they are extremely pertinent because the Regress Argument is implicitly committed to there not being any arguments for LOT beyond those that address the explananda it explicitly mentions. Seen in its proper light, then, the entire force of the Regress Argument depends upon there being arguments against the full range of positive reasons for endorsing LOT. Only then would the Regress Argument have any force, and in that case it probably wouldn't be necessary anyway. Far from having 'destroyed' LOT, the Regress Argument offers little or no reason to be sceptical of the hypothesis: the unsupported presupposition of the argument is just the one on which the whole issue really turns.[4]

References

Blackburn, S. 1984. *Spreading the Word*. Oxford: Oxford University Press.

Churchland,P.1986. *Neurophilosophy*. Cambridge, Mass: MIT Press.

Crane, T. 1995. *The Mechanical Mind*. London: Penguin.

Fodor, J. 1975. *The Language of Thought*. Cambridge, Mass: Harvard University Press.

Fodor, J. 1987. *Psychosemantics*. Cambridge, Mass: MIT Press.

Levelt,P.1993. Lexical access in speech production. In *Knowledge and Language*, ed. E. Reuland and W. Abraham. Dordrecht: Kluwer.

---

[4] Authors names are listed alphabetically.