You don't Know How you Think: Introspection and Language of Thought

Edouard Machery

You don't Know How You Think:

Introspection and Language of Thought

Edouard Machery

Abstract

The question, 'Is cognition linguistic?' divides recent cognitive theories into two antagonistic groups. Sententialists claim that we think in some language, while advocates of non linguistic views of cognition deny this claim. The Introspective Argument for Sententialism is one of the most appealing arguments for sententialism. In substance, it claims that the introspective fact of inner speech provides strong evidence that our thoughts are linguistic. This article challenges this argument. I claim that the Introspective Argument for Sententialism confuses the content of our thoughts with their vehicles: while sententialism is a thesis about the vehicles of our thoughts, inner speech sentences are the content of auditory or articulatory images. The rebuttal of the introspective argument for sententialism is shown to have a general significance in cognitive science: Introspection does not tell us how we think.

1 The Problem

The question, 'Is cognition linguistic?', divides recent cognitive theories into two antagonistic groups. Among each group, there are for sure some interesting and deep differences. Still, this

opposition is crucial. Several views of cognition, for example the recent dynamic systems approach (e.g., Thelen and Smith [1994]; Port and van Gelder [1995]) or the new wave of empiricism (Barsalou [1999], Prinz [2002]), have been explicitly developed to supersede sententialism; similarly, its adequacy has been the object of a passionate controversy in psychology, philosophy and artificial intelligence (e.g., McClelland et al. [1986]; Smolensky [1988], [1991]; Fodor and Pylyshyn [1988]).

For present purposes, *the sententialist hypothesis* claims that the mental representations (MRs) that feature in the cognitive processes constitute a *language*. According to *the Mentalese hypothesis*, the cognitive processes are defined over mental representations that constitute a *non natural language* (Fodor [1975], [1998a]; Jackendoff [1989]). To put it simply, when you entertain a thought that p, you token a sentence of a language that does not possess the characteristic properties of natural languages. On the contrary, some claim that human beings' cognitive processes are defined over expressions of *natural languages* (Sellars [1956]; Carruthers [1996], [1998a], [1998b], [2002]; Devitt and Sterelny [1999]). When you, English speaker, entertain a thought that p, you token a sentence of English that means p.

The non linguistic conception of cognition is really a mixed bag. It includes very different conceptions of cognition, which are only unified by their rejection of the sententialist orthodoxy.

Sententialists have put forward many arguments for their pet view of cognition (e.g., Fodor [1975]; Fodor and Pylyshyn [1988]; Horgan and Tienson [1996]; Carruthers [1998a], [1998b], [2002]). Among all those more or less technical arguments, one of them is particularly seductive. For it is based on our own experience, on the way most of us experience our own conscious thoughts.

> *In substance, the argument infers inductively from the introspective fact of inner speech*[1] *that our conscious propositional thoughts are natural language sentences*.[2]

When we introspect our own conscious propositional thoughts, we have access to thoughts expressed in natural language sentences; in other words, we find ourselves engaged in inner speech. This introspective fact is treated as evidence that we do think consciously in a natural language: as Carruthers states it ([1996], p. 50), 'introspection informs us, in fact, that many of our thoughts are expressed in natural language.'[3] I call this inductive inference the Introspective Argument for Sententialism, or, for short, the IAS.

In this paper, I purport to refute that claim:

The introspective fact of inner speech does not count as evidence that our conscious thoughts are expressed in a natural language.

In a nutshell, I want to push the following line of thought against the IAS: *the nature of the sententialist hypothesis implies that the introspective fact of inner speech cannot be evidence that we think in a language*. True, we know introspectively that we utter sentences in inner speech. However, introspection falls short of providing any evidence about the properties that mental representations have to possess for any form of sententialism to be true – or so I claim.

Before going to work, one important caveat. *I am <u>not</u> arguing against sententialism*, which may be true in general or, maybe, of some aspects of the cognition. Nor am I arguing against the idea that our conscious thoughts are natural language sentences. I am *exclusively* concerned with the introspective argument that is taken to support this view. I am claiming that this argument does not go through.

In the first section of this paper, I present in detail the introspective argument for sententialism. In the second section, I present my own objection and argue for its premises. The basic significance of sententialism is reviewed and the core of my objection is explained:

inner speech sentences are represented by auditory or articulatory images. In the last section, I rebut three objections to my argument. I conclude that the IAS does not go through and I emphasize the general significance of this conclusion for cognitive science.

2 The Introspective Argument for Sententialism

In the first section, I present and develop the IAS. The argument is the following:

1. When we introspect our conscious propositional thoughts, we find ourselves uttering natural language sentences, that is, we find ourselves engaged in inner speech.
2. If our theories of consciousness don't imply that introspection is systematically misleading, the best explanation of inner speech is that we think in a natural language.
3. Our best theories of consciousness don't imply that introspection is systematically misleading.
4. Hence, our conscious propositional thoughts (qua tokens) are natural language sentences.
5. Hence, sententialism is at least true of our conscious propositional thoughts.

Many philosophers have put forward this argument in a more or less developed form. For example, Devitt and Sterelny ([1999], p. 142), refers to it[4] :

> The hypothesis also has some introspective support. Speech often *seems* to be thinking aloud; thought often *seems* to be talking to oneself.

Particularly, this argument is the core of P. Carruthers' book, called *Language, Thought and Consciousness* ([1996]). He writes ([1996], p. 228):

> The best explanation of the available introspective data is that we mostly think (when our thinking is conscious) by imaging sentences of a natural language, and trains of thought consist of manipulations and sequences of such images.

In this section, I will be mostly concerned with Carruthers' formulation of the argument, for it is the most developed.

What about the premises of the argument? First, I am willing to grant premise 3.[5] For I believe that despite the intense work on consciousness, and arguably, the progresses that have been made, we are far from having the good theory of consciousness that would be required to evaluate seriously premise 3. Moreover, as we will see below, I focus instead on the second premise: I want to show that even if we grant that introspection is not misleading, the introspective fact of inner speech can't support sententialism.

What about the first premise? It is based on the introspective fact of inner speech. The idea is the following. Whenever we have an introspective access to a conscious propositional thought that p, this thought is expressed in a natural language sentence that means p. Whenever I introspect my conscious propositional thoughts, I find myself engaged in inner speech. Introspection is not supposed to suggest that all our conscious thoughts are natural language sentences: we have an introspective access to non propositional, non linguistic conscious thoughts, for example visual images. But premise 1 denies that we have any introspective experience of conscious propositional thoughts that are not expressed by natural language sentences.

Does our *phenomenology* support that premise? The fact of inner speech seems widespread. When we reason for ourselves about a problem, either theoretical or practical, our arguments seem often to be expressed in sentences. Or so it seems to me and to Carruthers...

Moreover, inner speech is not dedicated to serious matters. It seems to be present in all kinds of conscious thinking, from the idlest to the most serious. Now that does not show that we do not have any introspective access to propositional thoughts that are not expressed by natural language sentences. As far as I am concerned, the answer is quite moot.

Beside his own introspection, Carruthers refers to some *empirical data* collected by Hurlburt ([1990], [1993]). The experimental procedure is the following: subjects wear headphones through which they hear signals at irregular intervals. The instruction is to register their conscious thoughts at the time of the signal. *All normal subjects report the occurrence of inner speech on some occasions* (the minimum is 7%, the maximum 80%, more than half the subject report inner speech on more than half the occasions). Subjects report also emotions, visual images, and, finally, *wordless thoughts*.[6]

This last type of introspective thoughts is problematic for Carruthers, for it suggests that some people have an introspective access to conscious propositional thoughts that are not expressed by natural language sentences. Carruthers explains them away as routine self-interpretations: these subjects don't have any non-linguistic, conscious propositional thought, but they *ascribe* to themselves propositional thoughts in the absence of any sentence.[7] Since the self-ascription is fast and automatic, it may elicit the erroneous belief that one entertains non-linguistic conscious propositional thoughts. One may object that Carruthers seems to choose among the introspective data those that support his argument. Thus, his reply may look like *ad hoc*. Maybe. But Carruthers could plausibly reply that this kind of dialectic between theory and evidence is a kosher practice in science. In any case, I am willing to grant also the first premise for the sake of argument.

What about premise 2? It is supposed to be intuitive. The idea is plausibly the following. Introspection is supposed to be a possible, though not necessarily conclusive,

source of evidence about our conscious thoughts. And clearly, if a source of evidence provides the evidence that p, then, *ceteris paribus*, we ought to believe that p. Since we know introspectively that our conscious propositional thought tokens are natural language sentences (premise 1), it is justified to infer inductively that to think consciously that p is really to utter in inner speech a sentence that means p.

I claim nonetheless that premise 2 is wanting. More precisely, I endorse the Blindness of Introspection Thesis:

(BI) The introspective fact of inner speech cannot be evidence that our conscious thoughts are linguistic.

I am not claiming that introspection is never a source of evidence of any sort. It may provide some evidence for some theses. I am not asserting either that, as far as sententialism is concerned, inner speech cannot be a source of evidence at all.[8] But I claim that the fact of inner speech does not provide any evidence that sententialism is true of our conscious thoughts. This properly denies what the IAS claims.[9] In the next section, I defend the BI thesis.

3 The Argument for the Blindness of Introspection Thesis

The argument for the BI thesis rests on two premises:

A. Sententialism concerns the vehicles of our thoughts.

B. We do not have any access to the properties of the vehicles of our thoughts.

C. Hence, the introspective fact of inner speech cannot be evidence that our conscious thoughts are linguistic (BI).

I examine first premise A. It articulates the significance of any form of sententialism, included the natural language hypothesis. I have characterized sententialism as the claim that

the mental representations constitute a language. What does that mean? I suppose that a propositional attitude, a belief, a desire etc. is a relation between a subject and a proposition (I put aside our indexical beliefs). The sententialist hypothesis is the claim that the proposition is expressed by a mental state which consists of a linguistic symbol. In turn, a set of symbols is linguistic if i/ this set consists of atomic and complex symbols; ii/ the atomic symbols are ultimately the constituents of the complex symbols according to some recursive rules of formation of complex symbols; and iii/ the semantic properties of a complex symbol are recursively determined from the semantic properties of its atomic symbols and their mode of composition (Fodor and Pylyshyn [1988], p. 13). [10]

I want to insist on the second point, namely *the constituency principle*: a complex symbol is compounded out of other symbols.[11] Fodor ([1987], p. 137), points out that this principle is the core of the Mentalese hypothesis:

> The LOT story amounts to the claims that (1) (some) mental formulas have mental formulas as parts; and (2) the parts are 'transportable': the same parts can appear in *lots* of mental formulas.

This is really *the core of the sententialist hypothesis in general*: the constituency principle distinguishes the sententialist hypothesis from other types of intentional realism. Intentional realists admit the reality of mental states endowed with semantic properties: for example, they may endorse the characterization of a belief as a relation between a subject and a proposition. But they avoid any commitment about the nature of these mental states. On the contrary, what distinguishes sententialists is their willingness to make a specific claim, i.e., the constituency principle, about the nature of the mental states that express these propositions.

Now, the constituency principle concerns *the vehicles of the mental representations* and not their contents.[12] It does not state that some representations have a complex content, but that some representations have other representations as constituents. To claim that some

representations have a complex content does not distinguish the sententialist hypothesis from other types of intentional realism. Indeed, this is common ground among all intentional realists. And the complexity of the content is noncommittal on the nature of these mental states: a complex content can be expressed by any kind of mental representations, hence by a representation without any structure. For example, a flag can express the proposition [the way is free]. Hence, the constituency is a structural property of the vehicles of complex mental representations. Thus, sententialism is a thesis about the vehicles of our thoughts.

Premise B is really the core of the argument for the BI thesis. It is supported by a simple argument. This argument applies to our conscious propositional thoughts a principle that is widely accepted for our conscious non propositional thoughts. The principle is the following:

> *The fact that P is a property represented by a thought does not licence per se the inference that the thought itself possesses P.*

We are not thinking consciously only propositional thoughts. Most of us entertain conscious feelings and conscious images. Let us focus on visual images. Nowadays, not many people are willing to infer any property of the vehicles of these visual images from the properties of their content. For example, it is widely recognized that the visual image of a red apple does not have to be red: true, the visual image represents the property *red*, but that fact does not support the claim that the vehicle of this image also has that property. This is generally true of the properties that are represented by visual images.[13]

Now, the crucial point is the following: *the sentences that are uttered in inner speech are part of the content of our conscious thoughts*. When I say to myself the sentence 'Barcelona is a wonderful city', this sentence token is *represented* by a conscious thought.

And to be at the same time more precise and more audacious: the inner speech sentences are the content of mental representations that are produced by the auditory or articulatory *imagination*.[14] In short, inner speech sentences are the content of (auditory or articulatory) images. They are imagined.

This point is supported by various considerations. First, by some *phenomenological* evidence. Inner speech sentences possess the characteristic *detailed* phenomenology of conscious images. In particular, an inner speech sentence can be said in a particular way, with a particular accent, a specific speed and so on. Or so it seems introspectively. Similarly, we are able to imagine particular shades of colours or particular shapes (for example, irregular circles). Moreover, inner speech sentences are phenomenologically similar to auditory images, for example tunes that we sing for ourselves.

It is also supported by some *empirical* evidence. The study of auditory imagery is much less developed than the study of visual imagery. Still, some preliminary results suggest that inner speech and the imagination of someone else's utterances share some cognitive processes. Smith et al. ([1995]) have studied the effect of blocking the subarticulation (by repeating irrelevant sounds) on the imagination of someone else's utterances: the performance is significantly impaired. Crucially, in similar conditions, inner speech is also similarly impaired (Reisberg et al. [1991]). Moreover, recent brain-imaging studies suggest that inner speech and auditory verbal imagery (imagining hearing someone's voice) share a common component of verbal working memory and covert generation/articulation (McGuire et al. [1996]; Shergill et al. [2002]). This suggests that inner speech is a form of verbal imagination: I imagine hearing a sentence that I attribute to myself.[15]

The conclusion is straightforward: the fact that we use sentences in inner speech does not support the claim that our conscious thought satisfies the constituency structure. True, the

object thought about, namely the sentence, possesses the property of constituency, for a sentence is compounded out of words; but this does not licence any inference that the thought itself possesses the constituency property. Similarly, an imagined square is compounded out of four sides. And again, an imagined three-dimensional body, for example a human being, is compounded out of simple volumes. But no one would infer that the visual image of a square has four sides because the imagined square has that structure. Again, no one would infer that the visual image of a human being is compounded out of volumes because its content has that structure. For it is recognized that the nature of the contents of those images alone does not support any inference about the vehicles of those images. Similarly, the introspective access to inner speech does not support the claim that the vehicles of our thoughts are also sentences.

In short, the sentences that are uttered in inner speech are represented by conscious images. They are part of the content of these thoughts. Now, generally, the fact that the content of some thoughts possesses a property P does not licence *per se* the attribution of P to their vehicles. Hence, the linguistic nature of the sentences uttered in inner speech cannot be attributed to the thoughts themselves. We should accept premise B: We do not have any access to the properties of the vehicles of our thoughts. When are experiencing inner speech sentences, we represent those sentences, without having any access to the properties of their vehicles.

Conclusion. The blindness of introspection thesis claims that the introspective fact of inner speech cannot be evidence that we are thinking in a natural language. If sententialism is a claim about the vehicles of our thoughts, namely that they satisfy the constituency principle (premise A) and if inner speech sentences are represented by conscious, auditory or articulatory images (premise B) we have to endorse the blindness of introspection thesis. Now

if this thesis is granted, it is not true that the introspective fact of inner speech supports the claim that our conscious thoughts are linguistic (premise 2 of the IAS). Hence, the introspective argument for sententialism does not go through. The introspective experience of inner speech does not support the claim that our conscious thoughts are linguistic.

4 Objections and Replies

In the last section, I address three possible objections.

First, one could simply claim that *we do have an access to the vehicles of our conscious thoughts*. For we have the introspective impression that we know not only what we are thinking, but also *how* we are thinking what we are thinking (see, for example, Carruthers [1996], p. 207). I reply that this objection rests on an ambiguity: in one sense, it is true that we know how we are thinking consciously; but in another sense, which is the only relevant for sententialism, this claim is not true.

I have conceded that we have the introspective impression to think consciously in a natural language (premise 1 of the IAS). And it may be true that whenever we entertain a conscious thought whose content is the proposition [Barcelona is a wonderful city], we utter a natural language sentence that means that Barcelona is a wonderful city. For the sake of argument, I am even ready to concede that to think consciously that p, we have to entertain a natural language sentence that means p. Still, this sentence is *represented* by a conscious image. And this fact does not support the claim that the vehicles of our thoughts satisfy the constituency principle.

Hence, *in one sense*, it may be true that we know introspectively how we entertain consciously a thought that p: we may have to imagine a sentence that means p in order to think consciously that p. But that does not provide any evidence that the vehicles of our conscious thoughts are syntactically structured, which is the very point at stake in

sententialism. Hence, *in the relevant sense*, we do not know introspectively how we entertain a conscious thought that p.

An analogy can drive that point home. Consider a visual image of the map of Paris' subway. The map consists of lines that represent Paris' subways and of dots with names like 'Chatelet' or 'Gare du nord' that represent Paris' subway stations. The content of that visual image is itself a representation, as is the content of an auditory or articulatory image of a natural language sentence. Suppose that to decide how to go from one place to another, I have to imagine a map of the subway. I am thinking about Paris' subway by imagining its map. Clearly, *in one sense*, I know *how* I am thinking about Paris' subway: to think about Paris' subway in this orientation problem, I have to imagine its map. But, *in another sense*, I don't know introspectively how I am thinking about Paris' subway: that is, I do not have any access to the properties of the vehicle of this visual image. For the vehicle of the visual image of the map does not have to exhibit the spatial properties of the map: for example, it could be a description in Mentalese, as could be the vehicles of all visual images (Pylyshyn [2002], [2003]). In other words, the visual image does not provide any evidence that the vehicles of our thoughts about Paris' metro in this problem-solving task possess any property of the map, especially its spatial properties. Even visual images enthusiasts like Kosslyn recognize that visual images don't exhibit literally the spatial properties that they represent (e.g., Denis and Kosslyn [1999]).

Hence, I conclude that in some sense, we may know how we entertain consciously the thought that p: we may have to utter in inner speech a sentence that means p to think consciously that p. Similarly, we may know how we think about some objects in some problem solving task: we imagine visually these objects. But that does not show that we have an introspective access to the vehicles of our thoughts – hence that we know that we are thinking consciously in a language or that we are occasionally thinking in images.

The second objection targets the premise B of the argument for the BI in a more sophisticated way. It can be formulated as follows:

α. The syntactic properties of inner speech sentences causally drive the reasoning.

β. We have an introspective access to these properties.

γ. Only properties of the vehicles of our MRs are causally efficient.

δ. Hence, we have an introspective access to at least some properties of the vehicles of our MRs.

The argument seems to be sound. What about its premises? The second premise can be taken for granted. I take the third premise for granted too – though it is clearly controversial. It rests on the idea that semantic properties are not causally efficient.[16] Hence the causal powers of a mental representation depend on the properties of its vehicles. What about the first premise? For Humean reasons, we don't have any direct introspective access to the causal properties of our conscious propositional thoughts. However, it seems relatively clear that when we say in inner speech a sentence whose form is $p \rightarrow q$ followed by a sentence whose form is p, then we say in inner speech a sentence whose form is q. A natural explanation of that fact is that the form of the inner speech sentences is causally efficient. Hence, we should endorse the conclusion δ and reject the premise B of the argument for the blindness of introspection.

But this argument does not go through, for its first premise is wanting. The syntactic properties of inner speech sentences are not causally efficient. The fact that our conscious reasoning obeys some formal rules is due to the fact that we *believe*, implicitly or explicitly, that we should obey these rules: we believe for example that it is logically legitimate to utter a sentence whose form is q after having uttered a sentence whose form is $p \rightarrow q$ and a sentence whose form is p. In other words, the conscious transition is to be explained in intentional terms: it rests on the content of a belief about how we should reason and not on the causal

efficiency of the syntactic properties of the inner speech sentences. Two arguments support that view. First, a syntactically ambiguous sentence can feature in a conscious inference: for example, I can infer from the inner speech sentences 'if God does not exist, everything is allowed and human beings are the measure of everything' and 'God does not exist' the sentence 'everything is allowed and human beings are the measure of everything'. Since the sentence is ambiguous, its syntactic properties fall short of explaining the reasoning. Second, and crucially, there is a plausible way to distinguish *empirically* the two hypotheses. If the syntactic properties of the sentences are causally efficient*, one's inferential dispositions should not be cognitively penetrable*. For a modification of our beliefs does not affect the syntactic properties of these sentences. On the contrary, if the syntactic properties of the inner speech sentences are not causally efficient, *a cognitive penetrability* of our inferential dispositions is to be expected. For, in that case, we utter in inner speech a sentence whose form is q after having uttered a sentence whose form is $p \rightarrow q$ and a sentence whose form is p, because we believe that we should reason that way. If we modify the content of our beliefs about how we should reason, then our patterns of conscious reasoning should be modified too. Are our patterns of reasoning cognitively penetrable? I don't have any experimental evidence to offer. But the following seems to be rather plausible. Suppose that a philosopher is deeply convinced of the truth of the intuitionist logic. Besides, she has been using this logic in her written and spoken reasoning for years. Probably this would affect her reasoning dispositions when she is engaged in inner speech. This suggests that the conscious reasoning is cognitively penetrable; hence, that the form of our imagined sentences are not causally efficient.

Again an analogy with visual imagery can drive the point home. In the visual imagery literature, it is often claimed that the spatial properties of the visual images or some analogue thereof, explain many empirical results (Kosslyn [1994]). Let us consider particularly the

well-known image-scanning phenomenon. The finding is that the time required by the subjects to go through the visual image of a map is a function of the imagined distance between the points (Kosslyn et al. [1978]). This paradigm has been used to infer some properties of the metrics of mental images, as well as some properties of the mind's eyes, for example its visual angle (Denis and Kosslyn [1999]). The idea is the following. We are licensed to attribute the spatial properties of the content of our visual images to the visual images themselves because these spatial properties affect causally our performance in the mental scanning task. Properties of the contents of images (e.g., their metrics) are attributed to the vehicles because of their alleged causal influence on our cognitive performances. Clearly, the problem is similar to the argument put forward above (α to δ). Now, Pylyshyn has suggested that the performance of the subject can be explained as follows (Pylyshyn [2002]). The subjects *simulate* that they are seeing the map. Since they *know* that the time required to go visually through a real map is a function of the distance between the departure point and the arrival point, the larger the imagined distance between these two points, the longer they scan imaginatively the imagined map. If this is true, the correlation between the imagined distance and the scanning time does not show that the physical properties of the imagined map are causally efficient, hence that they are properties of the vehicles of the visual images. How are we to distinguish between these two explanations? If Pylyshyn's *intentional* explanation is true, the scanning behaviour should be cognitively penetrable. He writes (Pylyshyn [2002], p. 161):

> How is it possible to tell whether certain imagery effects reflect the nature of the imagery architecture or the person's tacit knowledge? (...) One theoretically motivated diagnostic (...) is to test for the *cognitive penetrability* of the observations. This criterion is based on the assumption that if a particular pattern of observations arises because people are simulating a situation based on their tacit beliefs then if we alter

their beliefs or their assumptions about the task, say by varying the instructions, the pattern of observations may change accordingly, in ways that are rationally connected with the new beliefs.

Pylyshyn's experiments seem to show precisely that (Pylyshyn [2002]).

The same argument applies in the case of our inner speech and in the case of our visual imagery. Properties that are represented by inner speech sentences and by visual images may seem to drive our cognitive processes. This is however likely to be an illusion. Our (maybe implicit) beliefs drive really those cognitive processes. I conclude that we should not be led by this appearance: the syntactic properties of inner speech sentences and the spatial properties of visual images do not drive our cognitive processes. Hence, we do not have any introspective access to the properties of the vehicles of our thoughts. The Blindness of Introspection thesis stands: The introspective fact of inner speech is irrelevant for the nature of our thoughts.

The last objection recognizes that we don't have any direct access to the properties of the vehicles of our thoughts (vs. the first two objections). But it claims nonetheless that the introspective fact of inner speech shows that these vehicles are linguistic. The idea is the following. Suppose that when we think consciously the proposition [Barcelona is a wonderful city], we always imagine the sentence 'Barcelona is a wonderful city'. How is this egregious fact to be explained? How are we to explain the fact that I need to imagine the sentence 'Barcelona is a wonderful city' to entertain consciously the proposition it expresses? It may be claimed that the best explanation is that the image of the sentence is itself compounded out of the words of the sentence. If the vehicle of the conscious image is itself the sentence, no wonder that to think consciously that Barcelona is a wonderful city, I have to utter the sentence 'Barcelona is a wonderful city'. On the contrary, if I deny that the vehicles of my

conscious thoughts are composed out of the very words I represent, say out of 'Barcelona', 'wonderful' etc., how can I explain this fact?

There may be after all an introspective argument for sententialism. We think consciously that p by imagining a sentence that means p. The best explanation of that fact is that the mental images consist of the very sentences expressed in inner speech. Hence our conscious thoughts are syntactically structured.

I claim that the idea that the introspective impression to think in a language is best explained by the sententialist hypothesis (more precisely by its natural language variant) is *not justified*. The fact that our conscious thinking requires inner speech sentences (if it is a fact) can have various functional explanations that are noncommittal about the nature of the vehicles of our thoughts. For example, Jackendoff ([1996]) claims that inner speech makes it possible to pay attention to the abstract aspects of our thoughts. In other words, it allows for a kind of metarepresentational self-monitoring. If the evolutionary function of our conscious thought is to allow for complex cognitive processes requiring this self-monitoring, then our cognitive architecture could be such that our conscious thinking requires inner speech sentences.[17] Hence, there are some alternative explanations of the fact that we use inner speech sentences to think consciously (again, if it is a fact). These alternative explanations do not assume that the vehicles of our conscious thoughts satisfy the constituency principle. And the argument to the best explanation suggested above does not show that these explanations are dead ends. Thus, we can't assume that the best explanation of the correlation between conscious thinking and inner speech is that our conscious thoughts are linguistically structured.

5 Conclusion

The introspective argument is clearly right to assume that we have an introspective access to natural language sentences in inner speech. I entirely endorse that premise. I am even ready to accept, at least for the sake of argument, that to entertain a conscious thought that p, we have to utter a sentence that means p (hence, that natural languages play a crucial role in conscious propositional thinking). But it is incorrect to infer from this introspective fact that we think in a natural language. For the thesis that we think in a language concerns the vehicles of our thoughts, while the sentences are represented by conscious, auditory or articulatory images. And it is generally true that we cannot attribute a property to the vehicles of mental images because the content of these images possesses it. *Hence, the introspective argument for sententialism is misguided.*

Finally, the argument for the blindness of introspection illustrates an *important general moral*. I referred repeatedly to the visual imagery controversy. For a powerful argument for pictorialism and the introspective argument for sententialism rest on a common fallacy: our access to the content of our conscious thoughts, visual images or inner speech sentences, elicits the introspective impression that we know *how* we think. And this introspective experience seems to support both the pictorialist position in the visual imagery controversy and the sententialist position in the language of thought controversy. But, as the bottom line of this article should have made clear, this reasoning commits the *vehicle*/*content fallacy*: properties of the content of our thoughts are fallaciously attributed to their vehicles. Thus, the rebuttal of the introspective argument for sententialism has a general significance for cognitive science: INTROSPECTION DOES NOT TELL US HOW WE THINK.

Acknowledgements

I would like to thank two anonymous referees for their comments on an earlier version of this paper.

Max-Planck Institute for Human Development

Center for Adaptive Behavior and Cognition

Lentzeallee 94

14195 Berlin

Germany

machery@mpib-berlin.mpg.de

References

Barsalou, L.W. [1999]: 'Perceptual Symbol Systems', *Behavioral and Brain Sciences,* **22***,* pp. 577-609.

Berg, J. [1999]: 'Troubles with Neo-Notionalism', *Philosophia*, **27**, 3-4, pp. 459-481.

Carruthers, P. [1996]: *Language, Thought and Consciousness*, Cambridge: Cambridge University Press.

Carruthers, P. [1998a]: 'Thinking in Language? Evolution and a Modularist Hypothesis', in P. Carruthers and J. Boucher *(eds)*, 1998, *Language and Thought*, Cambridge: Cambridge University Press, pp. 94-119.

Carruthers, P. [1998b]: 'Conscious Thinking: Language or Elimination', *Mind and Language*, **3**, pp. 323-342.

Carruthers, P. [2002]: 'The Cognitive Functions of Language', *Behavioral and Brain Sciences*, **25**, 6, pp. 657-674.

Denis, M., and Kosslyn, S.M. [1999]: 'Scanning Visual Mental Images: A Window on the Mind', *Cahiers de Psychologie Cognitive / Current Psychology of Cognition*, **18**, 4, pp. 409-465.

Devitt, M., and Sterelny, K. [1999]: *Language and Reality*, *An Introduction to the Philosophy of Language*, 2nd ed., Cambridge, MA: MIT Press.

Fodor, J. A. [1975]: *The Language of Thought*, New York: Crowell.

Fodor, J. A. [1987]: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.

Fodor, J. A. [1998a]: *Concepts: Where Cognitive Science Went Wrong*, Oxford: Oxford University Press.

Fodor, J. A. [1998b]: 'Do We Think In Mentalese? Remarks on Some Arguments of Peter Carruthers', in J. A. Fodor, 1998, *In Critical Conditions, Polemical Essays on Cognitive Science and the philosophy of Mind*, Cambridge, MA: MIT Press, pp. 63-74.

Fodor, J. A., and Pylyshyn, Z. W. [1988], 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition*, **28**, 3-71.

Horgan, T., and Tienson, J. [1996]: *Connectionism and Philosophy of Psychology*, Cambridge, MA: MIT Press.

Hurlburt, R. [1990]: *Sampling Normal and Schizophrenic Experience,* New York: Plenum Press.

Hurlburt, R. [1993]: *Sampling Inner Experience with Disturbed Affect*, New York: Plenum Press.

Jackendoff, R. [1989]: 'What is a Concept, that a Person may Grasp it?', *Mind and Language*, **4**, pp. 68-102.

Jackendoff, R. [1996]: 'How Language Helps Us Think', *Pragmatics and Cognition*, **4**, 1, 1-34.

Kaplan, D. [1990]: 'Words', *Proceedings of the Aristotelian Society*, Supplementary Vol. **LXIV**, pp. 93-119.

Kosslyn, S. M. [1994]: *Image and Brain*, Cambridge, MA: MIT Press.

Kosslyn, S. M., Ball, T. M., and Reiser, B. J. [1978]: 'Visual Images Preserve Metric Spatial Information: Evidence from Studies of Image Scanning', *Journal of Experimental Psychology: Human perception and Performance*, **4**, pp. 46-60.

McClelland, J. L., Rumelheart, D. E., and the PDP Research Group [1986]: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Cambridge, MA: MIT Press.

McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S., and Frith C. D. [1996]: 'Functional Anatomy of Inner Speech and Auditory Verbal Imagery', *Psychological Medicine*, **26**, pp. 29-38.

McGuire, P. K., Robertson, D., Thacker, A., David, A. S., Kitson, N., Frackowiak, R. S. J., and Frith C. D. [1997]: 'Neural Correlates of Thinking in Sign Language', *Neuroreport, Cognitive Neuroscience and Neuropsychology*, **8**, pp. 695-698.

Millikan, R. G. [1993]: 'On Mentalese Orthography', in. B. Dahlom *(ed.)*, 1993, *Dennett and its Critics*, Blackwell Publisher, Oxford, pp. 97-123.

Partee, B. H. [1984]: 'Compositionality', in L. Landman and Veltman F. *(eds)*, 1984, *Varieties of Formal Semantics*, Dordrecht: Foris, pp. 281-311.

Pelletier, F. J. [1994]: 'The Principle of Semantic Compositionality', *Topoi*, **13**, pp. 11-24.

Port, R., and van Gelder, T. J. [1995]: *Mind as Motion: Explorations in the Dynamics of Cognition*, Cambridge, MA: MIT Press.

Prinz, J. J. [2002]: *Furnishing the Mind*, Cambridge, MA: MIT Press.

Pylyshyn, Z. W. [1984]: *Computation and Cognition: Toward a foundation for cognitive science*, Cambridge, MA: MIT Press.

Pylyshyn, Z. W. [2002]: 'Mental Imagery, In Search of a Theory', *Behavioral and Brain Sciences*, **25**, 2, pp. 157-182.

Pylyshyn, Z. W. [2003]: *Seeing and Visualizing: It's not What you Think*, Cambridge, MA: MIT Press.

Reisberg, D., Wilson, M., and Smith, J. D. [1991]: 'Auditory Imagery and Inner Speech', in R. Logie and M. Denis *(eds)*, 1991, *Mental Images in Human Cognition*, Amsterdam: Elsevier, pp. 59-81.

Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Murray, R. M., and McGuire, P. K. [2001]: 'A functional study of auditory verbal imagery', *Psychological Medicine*, **31**, pp. 241-253.

Sellars, W. [1956/1997], *Empiricism and the Philosophy of Mind*, Cambridge, MA: Harvard University Press.

Smith, J. D., Wilson, M., and Reisberg, D. [1995]: 'The Role of Subvocalization in Auditory Imagery', *Neuropsychologia*, **11**, pp. 1433-1454.

Smolensky, P. [1988]: 'On the Proper Treatment of Connectionism', *Behavioral and Brain Sciences*, **11**, pp. 1-74.

Smolensky, P. [1991]: 'Connectionism, Constituency, and the Language of Thought', in B. Loewer and G. Rey *(eds)*, 1991, *Meaning and Mind: Fodor and its Critics*, Oxford: Basil Blackwell, pp. 201-227.

Thelen E., and Smith, L. B. [1994]: *A Dynamical Systems Approach to the Development of Cognition and Action*, Cambridge, MA: MIT Press.

Wettstein, H. [1988]: 'Cognitive Significance Without Cognitive Content', *Mind*, **97**, pp. 1-28.

Wittgenstein, L. [1956]: *Philosophical Investigations*, G. E. M. Anscombe *(ed.)*, 1956, New York: Macmillan.

---

[1] In the psychological literature (e.g., Smith et al. [1995]), the word 'inner speech' is sometimes used in a restricted way: inner speech requires covert muscles movements. The

subvocalized rehearsal that does not require covert muscles movements does not count as inner speech, so conceived. I use this term in a *wider* sense: it includes all the cognitive processes that result in the phenomenology of heard sentences in the head.

[2] This argument does not establish that our entire cognition is linguistic. It supports *a restricted version* of sententialism: *our conscious thoughts constitute a natural language*.

[3] Carruthers has been using this argument to support the natural language view of cognition against the claim that all thoughts are expressed in Mentalese. But since the natural language hypothesis is a version of sententialism, this argument supports also the sententialism.

[4] See also Wittgenstein [1958], par. 152; Wettstein [1988], p. 10; Berg [1999], p. 466.

[5] But see Fodor [1998b].

[6] Hurlburt's paradigm raises many questions. Particularly, the fact that the subjects write down their thoughts may enhance the importance of inner speech: they may have the impression that they have been thinking in a natural language because they have to express linguistically what they were thinking.

[7] This move supposes the plausible, though not trivial thesis that introspection is different from self-interpretation.

[8] For example, one could run the following productivity argument. We know introspectively that we have the capacity to entertain an infinite number of conscious thoughts (introspective productivity premise). But only a linguistic system of representation can allow for such a capacity (non introspective premise). Hence, our conscious thoughts require a linguistic system of representation. This argument is different from the IAS, for in this argument, sententialism is not derived from the mere fact of inner speech.

[9] See Devitt and Sterelny's and Carruthers' quotations above.

[10] Some may find this characterization too strong. Particularly, the third requirement could be criticized, given that it is not obvious that the semantics of natural languages satisfies it (for a

discussion of the semantic compositionality of natural languages, see, e.g., Partee [1984]; Pelletier [1994]). Besides, the recursivity built in the second requirement may also be criticized. Clearly, the notion of language can be cashed out in many ways.

[11] Notice that the constituency principle does not boil down to the requirement i/. For a system of representations may have simple and complex symbols, although the former are not the components of the latter, as has been shown by Smolensky [1988], [1991].

[12] I endorse the vehicle/content distinction (see Kaplan [1989] and Millikan [1993] for a discussion of some aspects of this distinction).

[13] Here is a possible objection. In fact, P(the image is red|the image represents a red object) > P(the image is red). Nonetheless, we do not infer that the image is red because we have independent reasons to doubt this conclusion. But, everything else being equal, the fact that we entertain the visual image of a red object licences defeasibly the claim that our image is red – or so the objection goes. But this line of reasoning is mistaken. The vehicle of a conscious image of a red object can be *any* kind of representations. According to Pylyshyn ([2002]), the null hypothesis is that our conscious visual images involve the same form of representations as the other MRs.

[14] The nature of the imagination required by inner speech is an *empirical* question (Smith et al. [1995]). It is possible that different forms of inner speech require different types of imagination.

[15] It has been objected that having a conscious thought that p cannot be imagining that p, for imagining and thinking are two different attitudes. However, if this kind of a priori stipulation were taken at face value, we could close a priori the visual imagery debate: we simply could not think through visual images.

[16] For example, because they are not local, if semantic externalism is true; or because they are not kosher from a naturalistic point of view.

[17] That we entertain idle conscious thoughts is not inconsistent with this speculative claim. That the evolutionary function of our conscious thought is to allow for complex cognitive processes does not imply that our conscious thought is restricted to these complex cognitive processes.