

# Inner Speech, Natural Language, and the Modularity of the Mind

Axel Gelfert  
(Department of Philosophy  
National University of Singapore)  
[axel@gelfert.net](mailto:axel@gelfert.net)

**Abstract** Inner speech is a pervasive feature of our conscious mental lives. Yet its function and character remain an issue of philosophical debate. The present paper focuses on the relation between inner speech and natural language and on the cognitive functions that various contributors have ascribed to inner speech. In particular, it is argued that inner speech does not consist of bare, context-free internal presentations of sentential (or subsentential) content, but rather has an ineliminable perspectival element. The proposed model of inner speech, which characterizes inner speech as akin to the testimony of an inner interlocutor, accounts for this perspectival element and, it is argued, is explanatorily superior, insofar as it better explains, amongst other phenomena, the often condensed character of inner speech.

**Keywords** inner speech; natural language; modularity; inference; cognition.

## 1. Introduction

The phenomenon of inner speech has long been described as “one of the most difficult to investigate”<sup>1</sup>, yet its pervasiveness and introspective salience in conscious thought make it an interesting topic of philosophical analysis. Rather than focus on the relation between inner speech and the ‘language of thought’ as a putative

---

<sup>1</sup> Vygotsky, 1986, 226.

medium of thought, the present paper focuses on its relation to natural language and on the functions that various contributors to the debate have ascribed to inner speech. In particular, I argue that inner speech does not consist of bare (context-free), internal presentations of sentential (or subsentential) content, but rather has an ineliminably perspectival element, much as in the external case of communication with an interlocutor. Inner speech, in a nutshell, may best be described as the testimony of an ‘inner interlocutor’.

The rest of this paper is organized as follows. In Section 2, I characterize the phenomenon of inner speech and give a brief summary of the debate, including its historical origins. This is followed, in Section 3, by a discussion of the role of natural language in inner speech, which is often characterized as enabling second-order cognition – that is, the ability to attend to our own thoughts. The idea that inner speech and conscious thought are closely related can be generalized and extended, within a modular framework, by regarding intermodular integration as a key function of our language system; the theoretical implications of this are discussed in Section 4. Drawing on the preceding discussion, Section 5 provides a programmatic sketch of how the phenomenon of inner speech may be interpreted from within a modular framework and argues for the importance of context and perspective in inner speech. Rather than eliminating context, inner speech allows for a switching of perspectives, much like in the case of communication with external interlocutors. The paper ends with a brief Conclusion (Section 6), which summarizes the main points.

## **2. The phenomenon of inner speech**

Inner speech is a pervasive feature of our conscious mental lives. Few people would dispute that we can introspectively ascertain that we often use words in thinking without making them audible (through speech) or visible (through reading and writing). Beyond this initial starting point, however, there is little agreement as to the character and function of inner speech.<sup>2</sup>

---

<sup>2</sup> For a survey of the debate, see Vicente & Martínez-Manrique, 2011.

Lev Vygotsky (1896–1934), in his pioneering work on developmental psychology, considered inner speech to be the “internal reconstruction of an external operation”<sup>3</sup>, which emerges fairly late in the child’s cognitive development. In his *Thought and Language*, first published in Russian in 1934, Vygotsky notes the deep connection between concept acquisition, cognitive awareness, and language. Inner (self-directed) speech is a particular form of language use, which is derived from external – that is, other-directed – speech through a gradual process of internalization. Initially, speech is social and serves purely communicative functions as an external tool for social interaction. Over time, a child may employ speech in an egocentric way, by transferring socially acquired behavioural patterns to his or her “sphere of inner-personal psychic functions”<sup>4</sup>. Such self-directed talk, which in the beginning is voiced out aloud, is then further internalized and becomes (silent) inner speech. In addition, the process of internalization brings about syntactic changes. Thus, inner speech tends to be highly abbreviated and, according to Vygotsky, consists entirely of predicates: “It is as much a law of inner speech to omit subjects as it is a law of written speech to contain both subjects and predicates.”<sup>5</sup> Through a process of internalization, language as an external medium of communication and instruction is being transformed into a resource that the child can ‘tap into’, for example in problem-solving contexts.

Instead of analyzing inner speech from a developmental angle, one may alternatively begin by considering the phenomenology of inner speech in adult cognizers. Doing so reveals that not all inner speech is of the ‘abbreviated’ sort, but rather that we also engage in more complex inner talk – especially in the run-up to (outward-directed) linguistic tasks, as when we inwardly rehearse what we will eventually say out loud. As Fernando Martínez-Manrique and Agustín Vicente have recently emphasized, any prospective theory of inner speech should do justice to the following two claims about its occurrence and overall character:

---

<sup>3</sup> Vygotsky, 1978, 57.

<sup>4</sup> Vygotsky, 1986, 35.

<sup>5</sup> Vygotsky, 1986, 243.

(i) We have phenomenological acquaintance with our inner voice even when we are not rehearsing linguistic actions, and in many cases we do not experience phonological representations of sentences but dispersed linguistic items.

(ii) We can experience richer, more sentence-like inner speech, typically but not exclusively related to linguistic activities.<sup>6</sup>

There is no implication here that these two modes of inner speech are mutually exclusive (except perhaps at their most extreme), or that there can be no intermediate cases; clearly, the extent to which inner speech is ‘rich’ or ‘abbreviated’ is a matter of degree. Indeed, empirical studies of how inner speech is perceived and recalled suggest a multiplicity of functions and contexts, ranging from undirected ruminating or daydreaming to explicit, goal-directed uses in order “to plan tasks, remember self-motivate, solve problems, plan when to do specific tasks, [...] rehearse upcoming conversations, read, write or calculate, study, control emotions, determine what to wear, self-censor, [and] replay past conversations”.<sup>7</sup>

While there is no hard and fast distinction between these two types of inner speech, given that they differ primarily in their phenomenological salience, which is a matter of degree, it is nonetheless useful to be aware of them as representing different ends of a spectrum. This way, one can explore possible transitions, for example from deliberate sentence-like inner speech (which, on occasion, may be derived from external speech, which is recalled in memory or for the purpose of self-regulation) to more dispersed and attenuated forms of inner speech. The ability to gradually internalize external representations, which fits with the phenomenology of more or less explicit (and more or less sentence-like) inner speech, also has repercussions for the ‘internal/external’ divide in relation to the extended mind thesis. David Rumelhart et al. (1986) raise this issue as follows:

“We can be instructed to behave in a particular way. Responding to instructions in this way can be viewed simply as responding to

---

<sup>6</sup> Martínez-Manrique & Vicente, 2010, 143.

<sup>7</sup> Morin, Utzl, & Hamper, 2011, 1717.

some environmental event. We can also remember such an instruction and ‘tell ourselves’ what to do. We have, in this way, internalized the instruction. We believe that the process of following instructions is essentially the same whether we have told ourselves or have been told what to do. Thus even here we have a kind of internalization of an external representational format (i.e., language).”<sup>8</sup>

Similarly, one may wonder whether inner speech is merely a way of tapping into an external resource – i.e. language – or whether it leads to any profound changes in the (prior) internal organization of the cognitive apparatus itself. Daniel Dennett argues for the latter view when he explicitly includes talking to ourselves among the ways in which “[w]e build elaborate systems of mnemonic associations” and, over time, “refine our resources by incessant rehearsal and tinkering, turning our brains [...] into a huge structured network of competencies”; the principal “components of this technology for brain-manipulation are words”.<sup>9</sup> On Dennett’s account, external speech – especially as received (and subsequently rehearsed) by children – is akin to the input involved in programming a massive parallel computer; over time such external input becomes internalized as “semi-understood self-commentary”. Once again, the shift from more sentence-like linguistic representations to more attenuated and dispersed forms of inner speech is thus being invoked in the service of a specific theoretical proposal about the workings of the mind. Contrasting his own views with those of Dennett, Andy Clark argues for a less ‘transformational’ role of linguistic representations: “Where Dennett sees public language as effecting a profound but subtle re-organization of the brain itself, I am inclined to see it as in essence an external resource which complements – but does not profoundly alter – the brain’s own basic modes of representation and computation.”<sup>10</sup> Finally, there is the question (to be taken up in the next section) of whether the phenomenon of inner speech gives us any indication as to whether we ‘think in words’ – and, if so, whether we should accord

---

<sup>8</sup> Rumelhart et al., 1986, 47.

<sup>9</sup> Dennett, 1998, 292.

<sup>10</sup> Clark, 2011, 25; Clark’s paper was first published in 1998.

natural language any privileged role in our analysis of the processes that sustain inner speech and of its role in cognition.

### **3. A privileged role for natural language?**

Clark, as mentioned above, conceives of language as a ‘resource’ – that is, as an external tool for communication which is then co-opted to suit the purposes of internal cognitive processing: “Language stands revealed as a key resource by which we effectively redescribe our own thoughts in a format which makes them available for a variety of new operations and manipulations.”<sup>11</sup> Throughout, he makes clear that ‘language’ here is to be understood as ‘public language’: “Public language, I shall argue, is just such a tool – it is a species of external artefact whose current adaptive value is partially constituted by its role in re-shaping” our cognitive and computational abilities.<sup>12</sup> We might then say that, in the same way that public language extends our cognitive abilities, so, by the same token, public language extends into our mental lives: inner speech is just such an extension of public language into the realm of our mind.<sup>13</sup> As Clark puts it, “inner rehearsals, when they occur, are quite literally models of linguistic production”<sup>14</sup> – that is, they may not only be construed as linguistic productions, from the standpoint of an external observer, but they literally are internalized natural-language items.

A number of authors, Clark included, regard it as a crucial feature of natural languages that, in addition to their obvious communicative function, they also allow for the fixation of our own thoughts, thereby

---

<sup>11</sup> Clark, 2011, 34.

<sup>12</sup> Clark, 2011, 21.

<sup>13</sup> Of course, on the extended mind hypothesis, language and other tools outside the skull are *already* constitutive of the mind, so that talk of public language ‘extending into’ our mental lives is strictly speaking redundant. Then again, the same applies to the notion of the ‘extended mind’ itself – which is why talk of ‘extendedness’ is best understood by way of contrast with more traditional conceptions that privilege intracranial mental processes. I would like to thank an anonymous referee for bringing this to my attention.

<sup>14</sup> Clark, 2011, 26.

imbuing them with a degree of permanence that transcends the fleeting character of the (non-linguistic) part of our mental lives. (See e.g. Clark, 2011, 34, and the critical discussion in Section 5 below.) Ray Jackendoff notes the “phenomenological difference” involved in linguistically representing thoughts to ourselves: without the availability of a linguistic medium, “you couldn’t experience your thoughts as linguistic images … a whole modality of experience would be simply absent – a modality which, as pointed out earlier, is very important to human experience”.<sup>15</sup> More than an experiential loss is at stake here: for, as Jackendoff notes, “[o]nly by having a linguistic modality is it possible to experience the steps of any sort of abstract thought”.<sup>16</sup> The ability to represent thoughts to ourselves in inner speech is thus regarded as a precondition for explicit logical reasoning. To be sure, some high-functioning animals that lack language, e.g. a monkey, may nonetheless engage in thought-processes that we can reconstruct as inferences, but although “the monkey has the thoughts, [...] she doesn’t hear the corresponding sentences in her head, because she has no linguistic medium in which to express them”.<sup>17</sup> The availability of a linguistic medium thus affords what Clark calls second-order cognition<sup>18</sup> – the ability to attend to our own thoughts.

While it is natural to read Jackendoff and Clark as identifying natural languages with the linguistic medium required for second-order cognition – quite explicitly, as mentioned earlier, in the case of Clark (2011) – more needs to be said about what the specific contribution of natural language to this process is. José Luis Bermúdez, in his *Thinking Without Words* (2003), provides an elaboration of this problem and draws a clearer distinction between the medium of thought and the specific linguistic character of inner speech. As Bermúdez sees it, thought in general is not limited to natural language as its medium; for example, there are types of problems “we solve by manipulating mental images and exercising the

<sup>15</sup> Jackendoff, 1996, 18.

<sup>16</sup> Jackendoff, 1996, 19.

<sup>17</sup> Jackendoff, 1996, 18.

<sup>18</sup> Clark, 2011, 34.

visual imagination”<sup>19</sup>, all of which are properly construed as thought processes. Some thought processes are unconscious – as we may realize on those occasions where, having abandoned a complicated problem, the solution suddenly occurs to us, having been worked out ‘in the back of our mind’. There are, of course, also processes we are conscious of – e.g. bodily sensations, emotions, moods –, which do not qualify as thoughts. Yet, where the two domains – propositional thought and consciousness – overlap, thoughts are necessarily expressed in language, or so Bermúdez argues:

“When we are conscious of propositional thoughts we are conscious of imagined sentences. What we introspect when we introspect our propositional thoughts in the manner required for the processes of second-order cognitive dynamics is inner speech.”<sup>20</sup>

Contrasting the case of “public language sentences” (*ibid.*) with other purported vehicles of thought, notably mental models and mental maps, Bermúdez argues that “[b]y a process of elimination [...] we have reached the conclusion that thoughts can only be the objects of the type of reflexive thinking in which thoughts are the objects of thought if they have natural language vehicles”.<sup>21</sup> Since such reflexive thinking does occur, the natural-language character of thought has thereby been established, or so the argument goes. This conclusion, however, is a little hasty, given that there is ample empirical evidence (including the phenomenon of ‘verbal overshadowing’, to be discussed in the next section) that not only are natural-language vehicles not necessary for successful reasoning, but they can even impede performance on certain cognitive tasks. This is not to deny that verbalization can have a positive effect on other tasks, for example by focusing attention, improving executive control, or aiding the retrieval

---

<sup>19</sup> Bermúdez, 2003, 160.

<sup>20</sup> Bermúdez, 2003, 160.

<sup>21</sup> Bermúdez, 2003, 163, *italics added*; for the detailed comparison of public language sentences with mental models and mental maps, see Bermúdez, 2003, 160–163.

of task goals<sup>22</sup>; rather, it cautions against the wholesale assertion that all propositional thoughts “that we can consciously introspect [...] take the form of sentences in a public language”.<sup>23</sup>

#### 4. The modularity thesis

Another account that attaches great, though not exclusive, significance to the natural-language character of conscious thought is Peter Carruthers’s massive modularity thesis, which holds that the human mind consists entirely, or mainly, of mental modules, each of which is adapted to a relatively narrow, domain-specific class of cognitive tasks. (See Carruthers, 2006.) One challenge which a theory of mental modularity must meet is to account for certain important and distinctive features of human cognition, of which Carruthers identifies three main categories: “flexibility of content; creativity of content; and abductive inferences performed upon such contents.”<sup>24</sup>

Previously, Carruthers (1996) had argued for the central role of natural language as a vehicle of thought, as opposed to thought processes being carried out in a separate ‘language of thought’. Given that we introspectively encounter our occurrent thought processes in (linguistically codified) inner speech, the law of parsimony demands that we should also consider our latent and unconscious thoughts as being codified using natural-language resources, or so the argument went. Carruthers subsequently relaxed this demand, insofar as he shifted his attention from questions concerning the nature of the vehicles of conscious thought to the function of language as tool for thinking. One of the challenges that emerges from the purported massive modularity of the mind is how to account for the remarkable degree of integration across different modules. If the mind indeed consists of an assemblage of distinct modules, each of which is adapted to the demands of a highly specific demands and largely independent of other modules, the question arises what can hold such

---

<sup>22</sup> See for example Miyake et al., 2004.

<sup>23</sup> Bermúdez, 2003, 159–160.

<sup>24</sup> Carruthers, 2003, 503.

an assemblage together and enable the integration of information coming from the various modules. Language is a prime candidate for the task of achieving intermodular integration, given that it is already in the business of integrating information – outside information, that is – and making it accessible to other modules for cognitive processing. Even without delving into the details of how the language system works, we can see that its dual character as an input and output system imbues it with precisely the features that are required for achieving integration and information exchange across different modules. Carruthers goes even further by suggesting that the language system can globally ‘broadcast’ linguistically represented information to the other modules, thereby placing it at the centre of the human mind’s cognitive activity.

In his most recent work, Carruthers (2011, 2015) develops his theory of the mind further by arguing that all conscious thought is sensory-based (i.e., involves visual-imagistic, proprioceptive, or auditory imagery, including inner speech), while any “amodal” (i.e., non-sensory) propositional attitudes must necessarily be unconscious. While the latter can be causally operative, only the former can make it into consciousness, via a dedicated “special-purpose working memory system”. This means that, in order to acquire knowledge of our own minds, which after all includes such items as goals, intentions, desires, and propositional attitudes such as belief of various sorts, we cannot count on any form of direct privileged access; instead, “we have to turn our mindreading capacities on ourselves, drawing inferences from sensorily accessible cues (including not only our own overt behavior and circumstances, but also such things as our own inner speech and visual imagery)”.<sup>25</sup> Thus, to paraphrase one of Carruthers’s examples, we may take the unconscious decision to leave for the bus, which subsequently causes us, in combination with other factors, to rehearse in inner speech the sentence “I’ll leave for the bus now”: “This is globally broadcast and received as input by language-comprehension and mindreading systems, leading it to be heard as a decision to leave for the bus now.” Yet, we should resist the idea that the original decision “has somehow become bound into the content of a sentence

---

<sup>25</sup> Carruthers, 2015, ix.

in inner speech” (thereby rendering the decision itself conscious in some way) since, Carruthers argues, the sentence embeds not the decision itself, but a “higher-order judgment to the effect that one is taking a decision”.<sup>26</sup>

Inner speech, then, is an especially effective way of broadcasting content to various subsystems, but it is not the exclusive basis of thought. Rather, it provides material for us to engage with in order to gain indirect and interpretive knowledge of our own minds, which is “no different in principle from our access to the thoughts of other people”.<sup>27</sup> If we are under the impression that we have privileged access to knowledge of our own minds, in a way that is different from how we gain knowledge of the thoughts of other people, then this is solely a reflection of the greater range and availability of contextual cues and background knowledge when bringing our mindreading capacities to bear on ourselves. Not only are “[m]any of the same contextual cues [...] available in the first-person as in the third”, but, unlike in the case of interpreting another person, we can also ascertain through introspection what we are attending to. On this account, our mindreading capacities may themselves be thought of “as a consumer of global broadcasts”, with the latter often (but not always) taking the form of inner speech; mindreading, then, is one of the beneficiaries of Carruthers’s hypothesized “global broadcast architecture” of the mind.<sup>28</sup>

Integration via the broadcasting of content across mental modules, of course, is not limited to humans, but also occurs in non-linguistic (= non-human) animals, so it must be phylogenetically prior to the emergence of language. Indeed, for Carruthers, the non-linguistic mechanism for intermodular integration provides a blueprint for the (phylogenetically later) language-based mechanism, in spite of important differences.<sup>29</sup> How, then, can some thinking be globally broadcast, in such a way that it registers at a conscious level and is

---

<sup>26</sup> Carruthers, 2015, 98.

<sup>27</sup> Carruthers, 2015, ix.

<sup>28</sup> Carruthers, 2011, 51.

<sup>29</sup> I am here following the very concise summary of Carruthers’s account found in Martínez-Manrique & Vicente, 2010, 155.

made accessible to other modules? Roughly speaking, a course of action can be terminated just before being executed, but after quasi-perceptual representations of it have been formed, which may then be sustained through the attention of executive systems or through the attenuated activity of motor systems.<sup>30</sup> Such action-schemata may then be globally broadcast across the whole architecture of mental modules, serving as input for the conceptual modules. This allows for the ‘real-time’ representation of contemplated courses of action and for the evaluation of their consequences before a decision is taken. On this view, non-linguistic creatures, such as non-human animals, are able to (non-linguistically) evaluate different courses of action, but only by, in some sense, mimicking them and assessing, in a quasi-perceptual way, the imminent consequences that would have occurred, had the action been seen through to completion.

Humans, in virtue of language, have an additional pathway for broadcasting information across modules. Rather than quasi-perceptually imagining (non-linguistic) actions and their consequences directly, in the way just described, we can also engage in imagined speaking; hence, the salient role of inner speech in the mental lives of human reasoners. In doing so, we send the requisite instructions to our language production module, yet abort the actual production of speech just in time (in much the same way as we would suppress the actual carrying out of the action in the non-linguistic case). As a result, speech is not uttered, but “quasi-produced”, in the form of phonological images – that is, as inner speech, which, in turn, is received and decoded by the input linguistic module, and is then broadcast to those modules in charge of extracting further information. As Martínez-Manrique and Vicente put it: “Our conscious inner talk thus consists in this rehearsal of linguistic actions.”<sup>31</sup>

Carruthers’s account of the mind as being marked by massive modularity, with the language module being accorded an important role in achieving cross-module integration, is ambitious and, it seems fair to say, somewhat speculative. Yet it has much going for it: for one, as discussed, it achieves some degree of theoretical unification

---

<sup>30</sup> See also Carruthers, 2011, 49.

<sup>31</sup> Martínez-Manrique & Vicente, 2010, 155.

between our accounts of animal minds and human minds – precisely (and perhaps ironically) by giving up on the unitary nature of the mind itself, conceiving of it as a vast collection of mental modules. And it is not all speculation: by committing himself to specific pathways of cognitive processing, Carruthers explicitly allows for, one might even say: invites, the empirical study of his account and its consequences. It is from this angle, however, that an empirical challenge has been mounted against the alleged central role of the language module in achieving seamless cross-module integration. The empirical phenomenon in question is known as verbal overshadowing, which occurs “when verbalizing mental contents deteriorates the performance of a task in which those contents appear to be involved”.<sup>32</sup> The effect was first described by Schooler and Engstler-Schooler (1990) who, in a series of experiments with putative witnesses to a crime (simulated in the experiment by showing participants a short video clip of a bank robber), demonstrated that the act of providing a verbal description leads to a considerable deterioration of a witness’s memory of the perpetrator, as shown by a decrease in the reliability of subsequently picking out the suspect’s face from a photo line-up. It is not the case that the subjects in the experiment are merely distracted, as might happen if they were being fed extraneous linguistic information while being asked to simultaneously carry out – now less successfully, it turns out – a non-linguistic task. Rather, in the case of verbal overshadowing the linguistic information is relevant to the task at hand and can even convey accurate information about the bank robber’s appearance, which one might have expected to aid the identification process, but this is not what one finds in the experiment. At first sight, this might be considered *prima facie* evidence against the general view, of which Carruthers’s account is an instance, that language serves as a lingua franca for a great number of mental operations.<sup>33</sup> Yet upon closer inspection, the theoretical repercussions of the phenomenon of verbal overshadowing are not quite as clear as this. For example, verbal overshadowing varies with levels of expertise (including familiarity

---

<sup>32</sup> Martínez-Manrique & Vicente, 2010, 157.

<sup>33</sup> This is how Martínez-Manrique & Vicente, 2010, interpret the matter.

with relevant vocabulary). Thus, in Melcher and Schooler's study (1996) of wine-tasting, experimental subjects with varying degrees of expertise ('novices', 'intermediates', and 'experts', respectively), after having had to give a verbal description of the flavour of a wine they had sampled, were asked to correctly re-identify it in a second round of tasting. Verbalization, it turned out, helped those considered 'novices', had no effect on designated 'experts', yet led to deterioration in the performance of 'intermediates'. This, as Giovanna Colombetti has argued, suggests that it is not verbalization per se that is a problem, but poor verbalization, which reflects a mismatch between the (borderline-expert) sensory capacities of intermediates and their impoverished vocabulary, which does not allow them to track their sensory experiences in a sufficiently fine-grained way.<sup>34</sup>

## 5. Inference, the 'inner interlocutor', and the perspectival character of thought

Though the empirical results indicate that not just any form of verbalization by the language module will give rise to cognitive integration (even if it involves relevant contents), all this shows is that the cognitive function of language is multifaceted. I shall return to this point below, where I will offer some thoughts on which factors may be relevant in such cases. For now, and in order to pave the way for the subsequent discussion, I shall turn to another aim of Carruthers's account: viz., to give an explanation of the flexibility of thought and of our ability to engage in abductive reasoning.

Carruthers links the evolution of our ability to engage in abductive reasoning to the emergence of language. Ontogenetically, too, one finds that early childhood development is marked by the gradual emergence of a capacity "to generate, and to reason with, novel suppositions or imaginary scenarios".<sup>35</sup> In children this ability is perhaps most evident in pretend play, yet it carries over into adulthood in the form of creative thinking. How could such a capacity

---

<sup>34</sup> See Colombetti, 2009.

<sup>35</sup> All quotes in this paragraph are from Carruthers, 2003, 511–512.

– “to suppose that something is the case (that the banana is a telephone; that the doll is alive), and then think and act within the scope of that supposition” – have arisen within a modular framework? As already noted by Vygotsky (see Section 2), much of the cognitive activity that accompanies childhood pretend play as well as creative supposition-generation in adults, takes the form of rehearsed inner speech, which, in Carruthers’s interpretation, serves the function of globally broadcasting contents across a number of mental modules. When, as a result of this process, contents are inserted into novel cognitive contexts, this can associatively give rise to new contents. Of course, we do not merely generate new contents for the sake of novelty, but we often do so in order to solve problems and in contexts that require action, and for these ends we must also “come to believe some of our suppositions”.

Deciding on the best among a number of possible solutions to a problem, or choosing the best from a set of imagined courses of action, engages the very cognitive capacities that are commonly seen to be at work in abductive reasoning. Inferring the most likely outcome of a proposed course of action involves entertaining, and mentally rehearsing, a number of hypotheses about what will (or might) happen. Typically, such scenarios and hypotheses are rehearsed through inner speech (although, of course, we also have at our disposal the non-linguistic mechanism of quasi-perceptual simulation); when this is so, i.e. when “the hypotheses in question are expressed in language, the problem of inferring the best explanation reduces to the problem of deciding which of the candidate sentences to believe in the circumstances”.<sup>36</sup> Echoing the original Vygotskian insight into the social origins of inner speech, Carruthers notes “that the principles of testimony-acceptance are historically and developmentally prior to the principles of inference to the best explanation”.<sup>37</sup> When viewed from this angle, the problem of deciding between candidate sentences or hypotheses we find ourselves presented with is, of course, a familiar one: it is a version of the problem of which testimony to accept, and when. In both cases, we

<sup>36</sup> Carruthers, 2006, 364.

<sup>37</sup> Carruthers, 2003, 514.

find ourselves confronted with candidate sentences, whether presented to us by an external interlocutor or emerging as inner speech from internal thought processes. And just as we do not accept every claim or utterance by an interlocutor and form a testimony-based belief on its basis, we do not accept every instance of inner speech as expressing a belief or as worthy of being relied upon in conscious deliberation – not least because we are aware of the varied functions of inner speech (see end of Section 3 above).

This parallel between linguistic hypothesis-generation and the case of testimony points to a fruitful basis for readjusting our paradigmatic way for thinking about inner speech, by modelling inner speech after the way we receive testimony. Just like the testimony of others, so inner speech, on the overall account of cognition endorsed here, is best treated as the product of more complex processes – in this case, largely unconscious thought processes – which are not readily transparent to its recipient. This also applies to the phenomenology associated with the reception of such linguistic items; indeed, while there is bound to be a rich phenomenology in any particular instance of someone who, on that occasion, encounters either an external utterance or an instance of inner speech, it is by no means clear that there is such a thing as a unitary phenomenology of encountering linguistic items, let alone distinct phenomenologies of understanding inner speech and understanding testimony, respectively.<sup>38</sup> As William Robinson notes, “[p]eople who understand what they are saying, or what is being said to them, are people who do not experience a problem with what is being said”.<sup>39</sup> When viewed from this angle, inner speech may be conceived of as the testimony of what one might call an inner interlocutor. Such a readjustment of our guiding conception of what inner speech is brings with it a new set of considerations that previously were not foregrounded. For example, real testimony is

---

<sup>38</sup> Perhaps because of this, few epistemologists of testimony nowadays argue that there is a distinct, identifiable phenomenology associated with testimonial acceptance or rejection. As Martin Kusch puts it, there “is no determinate phenomenology of testimony over and above imagined talk” (Kusch, 2002, 25). See also Gelfert, 2014, 64–68.

<sup>39</sup> Robinson, 2011, 201; *italics added*.

almost never context-free, but is accompanied by empirical circumstances, background knowledge, and other parameters (such as ‘conversational score’), all of which may inform the interpretation and contribute to, or detract from, the acceptability of the interlocutor’s testimony. In testimonial encounters, we are not usually confronted with ‘bare’ messages, whose context and causal ancestry is inscrutable to us, but instead with empirically rich conversational situations emerging from, often enough (though not always), familiar contexts.<sup>40</sup> All of this facilitates our assessment of individual instances of testimony, which would otherwise lack a justificatory basis. Yet, as we saw in our earlier discussion of Carruthers’s position (see Section 4), the cognizer who encounters an episode of inner speech can likewise draw on a rich array of contextual cues and background knowledge for its interpretation – and indeed must do so if he is to gain knowledge about his own mind from such episodes.

It may be promising, then, to model our engagement with inner speech as an encounter with the testimony of an ‘inner interlocutor’. At first sight, this might seem implausible: whereas in the case of testimonial knowledge there is a clear asymmetry between the speaker (who knows) and the recipient (who depends for his knowledge on the speaker), it might seem that in our case no such asymmetry exists. After all, what could my ‘inner interlocutor’ possibly tell me that I do not already know? This objection, however, is misguided, insofar as we are not dealing with two separate, fully formed epistemic subjects, but instead with relations between subpersonal mental modules. In this sense, the suggested parallel with the testimonial case may be considered partly metaphorical. But it is more than just metaphorical: what the parallel with testimony highlights is that there is more to inner speech than bare sentential (or subsentential) presentations of content. When we engage in inner speech, we adopt the perspective of

<sup>40</sup> Contemporary epistemology of testimony tends to exaggerate the inscrutability of testimonial encounters, by unduly decontextualizing cases of testimonial knowledge, for example by focussing on one-off encounters between strangers; on this point see Olmos, 2008, and Gelfert, 2014, 85–90. An extreme case would be Tyler Burge, 1993, who discusses testimony in terms of bare (context-free) presentations of intelligible messages.

a recipient of testimony – even if the testifier is simply our ‘inner interlocutor’ – and in doing so, we draw on whatever else is available to us in coming to an overall assessment or conclusion. As in the case of a recipient of (external) testimony, such assessment is subject to competing desiderata and will deploy various heuristics as part of a satisficing strategy from the recipient’s perspective.<sup>41</sup> There will, of course, be differences between how we consciously evaluate inner speech and how we assess external testimony: for example, external testimony is subject to many social conventions, the violation of which may give us cause to reject the testimony (or at least request clarification). Furthermore, it has been argued that certain criteria and maxims that are appropriate for external testimony, such as relevance-theoretic considerations of informativeness and economy, cannot be brought to bear on the case of inner speech.<sup>42</sup> To be sure, when it comes to inner speech, we cannot strictly distinguish between a speaker’s intention to communicate and a hearer’s recognition of such an intention, since we are dealing with aspects of one person’s inner mental life. Yet, over time, most of us have learnt to distinguish, and alternate, between different (typically context-dependent) perspectives of our inner interlocutor. Indeed, the ability to adopt different perspectives in interpreting inner speech has been deemed by psychologists to be “one of the main differences between a healthy person’s internal dialog and the pathology of hearing voices in mental illness”.<sup>43</sup> Conscious thought as encountered in inner speech, on this model, is ineliminably perspectival, and is marked further by a switching of perspectives in dialogue with our inner interlocutor – that is, in response to episodes of inner speech we experience as part of our conscious mental lives.<sup>44</sup>

---

<sup>41</sup> On this point, see Gelfert, 2010, 391–392.

<sup>42</sup> See García Murga, 1998, 78.

<sup>43</sup> Puchalska–Wasyl, 2015, 444.

<sup>44</sup> Martínez-Manrique and Vicente make a similar suggestion when they write that “precisely because the actual content of a thought takes into account perspective and context, while NL [natural-language] sentences do not, that having conscious thought does not amount to perceiving inner speech” (2010, 151).

It is worth contrasting this proposal with the context-free, aperspectival picture that has so far dominated philosophical discussions of the phenomenon of inner speech and its role in cognition. Here, for example, is Clark extolling the virtues of “the coding system of public language” for the purposes of inner speech:

“By ‘freezing’ our own thoughts in the memorable, context-resistant and modality-transcending format of a sentence we thus create a special kind of mental object – an object which is apt for scrutiny from multiple different cognitive angles, which is not doomed to alter or change every time we are exposed to new inputs or information, and which fixes the ideas at a fairly high level of abstraction from the idiosyncratic details of their proximal origins in sensory input.”<sup>45</sup>

Such a mental object, Clark continues<sup>46</sup>, is “ideally suited” for the “close and repeated inspections” that may be subsumed “under the rubric of attending to our own thoughts” – not least since the quoted characteristics give stability and permanence to our mental operations, thereby allowing for “self-inspection and self-criticism”. These are valid points, yet by asserting that conscious linguistic thought “minimizes contextuality”, Clark all but seems to deny the perspectival character of thought.<sup>47</sup> This near denial puts pressure on the very possibility of successfully discharging the functions of inner speech that Clark identifies as supremely important. Take the example of self-criticism and its complex phenomenology, which, however vexed they may be in detail, involve not the elimination, but the switching of perspectives and contexts. Robinson, describing an experience of stage fright while presenting a conference paper, gives vivid expression to the phenomenology of switching between different perspectives:

---

<sup>45</sup> Clark, 2011, 34.

<sup>46</sup> All snippets in this and the next sentence are from Clark, 2011, 34.

<sup>47</sup> In the interest of fairness, it should be noted that Clark weakens his claim in a footnote, where he speaks of a “*relative* context-independence of the signs and symbols of public language” (italics added), thereby shifting the focus away from sentences towards their constituent signs and symbols.

“[S]omewhere in the middle of [reading my paper], I asked myself how I was doing. There followed a strange few moments in which, while continuing to read my paper, I wondered whether my asking myself how I was doing was affecting my delivery, judged that it was not, opined that it soon would if I didn't stop asking these questions, and admonished myself to get back to concentrating on the subject matter of my paper. During this (fortunately brief) period, I was also aware that while the words were continuing to come just as planned, I had no idea what I was saying.”<sup>48</sup>

As this example illustrates, inner speech often involves distinguishing between the speaker and the recipient<sup>49</sup>; the recipient of inner speech – who, on this occasion, is himself engaged in giving external testimony to others – finds himself confronted with pronouncements on his (external) performance, before switching back to a state of immersion in his external course of actions. In the case of stage fright, this may lead to an overall deterioration of performance on the external task, but when it comes to mental activities such as planning, analyzing, or self-criticism, adopting – and finding oneself confronted with – different perspectives in inner speech, is arguably part and parcel of their proper functioning. Bare, internal presentations of natural-language content alone can hardly constitute such activities; in order for them to acquire their specific functions, they require ‘voice’ – that is, we need to treat them as the criticism, advice, or suggestions we receive from an inner interlocutor.

## 6. Conclusion

Inner speech is a complex phenomenon and its philosophical study is richly rewarding, given that it raises numerous questions at the intersection of the philosophy of language, mind, and cognition. In this short paper, I have argued for a reconsideration of the standard view of the character of inner speech. While inner speech is often

---

<sup>48</sup> Robinson, 2011, 200.

<sup>49</sup> On this point, see also Puchalska–Wasyl, 2015, 444.

regarded as the bare internal presentation of sentential content, this risks underestimating the significance of context and perspective, which render inner speech more akin to external communication than is standardly realized. The proposed model of inner speech, which likens inner speech to the testimony of an inner interlocutor, is able to account for its perspectival character as well as for certain empirical evidence concerning the character of inner speech, such as its often condensed character. The latter can be explained by analogy with the way we tailor utterances to external interlocutors: in situations where we have reason to assume significant overlap in background beliefs, we may employ more abbreviated ways of linguistic expression than in cases where background beliefs are not shared.<sup>50</sup> Similarly, when faced with novel or unusual situations, or when thinking about theoretical issues that require making one's theoretical assumptions explicit, we will typically engage in more complex, more expanded inner speech. Conceiving of inner speech as akin to testimony of an inner interlocutor, I submit, is phenomenologically and explanatorily superior to a view that strips conscious linguistic thought of all remnants of context and perspective.

## Acknowledgments

I am grateful to two anonymous referees for their helpful and constructive feedback.

## References

- Bermúdez, J.L., 2003, *Thinking Without Words*, New York, Oxford University Press.  
Burge, T., 1993, Content Preservation. *The Philosophical Review*, 102 (4), 457–488.

---

<sup>50</sup> See also Martínez-Manrique & Vicente, 2010, 156, on this point: “If inner talk is, in effect, talk to oneself [...], then we are in a position to explain those cases in which inner speech is condensed.”

- Carruthers, P., 1996, *Language, Thought, and Consciousness: An Essay in Philosophical Psychology*, Cambridge, Cambridge University Press.
- 2003, On Fodor's Problem. *Mind & Language* 18 (5), 502 – 523.
- 2006, *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*, Oxford, Oxford University Press.
- 2011, *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford, Oxford University Press.
- 2015, *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*, Oxford, Oxford University Press.
- Clark, A., 2011, Magic Words: How Language Augments Human Computation. In A. Clark, J. Toribio (eds.), *Language and Meaning in Cognitive Science: Cognitive Issues and Semantic Theory*, Abingdon, Routledge, 21–39.
- Colombetti, G., 2009, What Language Does to Feelings. *Journal of Consciousness Studies* 16 (9), 4 – 26.
- Dennett, D., 1998, Reflections on Language and Mind, In P. Carruther & J. Boucher (eds.), *Language and Thought: Interdisciplinary Themes*, Cambridge, Cambridge University Press 284 – 294.
- García Murga, F., 1998, Communication, Information, and Relevance. *Revista Alicantina de Estudios Ingleses*, 11, 75 – 83.
- Gelfert, A., 2010, Reconsidering the Role of Inference to the Best Explanation in the Epistemology of Testimony. *Studies in History and Philosophy of Science* 41 (4), 386 – 396.
- 2014, *A Critical Introduction to Testimony*, London, Bloomsbury.
- Jackendoff, R., 1996, How Language Helps Us Think. *Pragmatics & Cognition* 4 (1), 1 – 34.
- Kusch, M., 2002, *Knowledge by Agreement*, Oxford, Oxford University Press.
- Martínez-Manrique, F., Vicente, A., 2010, 'What the...!' The Role of Inner Speech in Conscious Thought. *Journal of Consciousness Studies* 17 (9–10), 141 – 167.
- Melcher, J.M. & Schooler, J.W., 1996, The Misremembrance of Wines Past: Verbal and Perceptual Expertise Differentially Mediate Verbal Overshadowing of Taste Memory. *The Journal of Memory and Language*, 35 (2), 231 – 245.
- Miyake, A., Emerson, M. J., Padilla, F., Ahn, J.-C., 2004, Inner Speech as a Retrieval Aid for Task Goals: The Effects of Cue Type and Articulatory Suppression in the Random Task Cuing Paradigm. *Acta Psychologica* 115 (2–3), pp. 123 – 142.
- Morin, A., Utzl, B., Hamper, B., 2011, Self-Reported Frequency, Content, and Functions of Inner Speech. *Procedia – Social and Behavioral Sciences* 30, 1714 – 1718.

- Olmos, P., 2008, Situated Practices of Testimony: A Rhetorical Approach. *Theoria: Revista de Teoría, Historia y Fundamentos de la Ciencia* 23 (1), 57 – 68.
- Puchalska-Wasyl, M. M., 2015, Self-Talk: Conversation With Oneself: On The Types of Internal Interlocutors. *The Journal of Psychology* 149 (5), 443– 460.
- Robinson, W., 2011, A Frugal View of Cognitive Phenomenology, In T. Bayne & M. Montague (ed.s), *Cognitive Phenomenology*, Oxford, Oxford University Press, 197 – 213.
- Rumelhart, D.E., Hinton, G.E., McClelland, J.L., 1986, A General Framework for Parallel Distributed Processing. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2: *Psychological and Biological Models*, eds. J.L. McClelland, D.E. Rumelhart, and the PDP Research Group), Cambridge (Mass.), The MIT Press, 45 – 76.
- Schooler, J.W., Engstler-Schooler, T.Y., 1990, Verbal Overshadowing of Visual Memories: Some Things are Better Left Unsaid. *Cognitive Psychology* 22 (1), 36 – 71.
- Sperber, D., Wilson, D., 1995, *Relevance: Communication and Cognition* (2nd ed.), Oxford, Blackwell.
- Vicente, A., Martínez Manrique, F., 2011, Inner Speech: Nature and Functions. *Philosophy Compass* 6 (3), 209 – 219.
- Vygotsky, L., 1978, *Mind in Society: The Development of Higher Psychological Processes* (ed. by M. Cole, V. John-Steiner, S. Scribner, E. Souberman), Cambridge (Mass.), Harvard University Press.
- Vygotsky, L., 1986, *Thought and Language* (ed. by A. Kozulin), Cambridge (Mass.), The MIT Press.