Francesco Fanti Rovetta

Chapter 14
Two Kinds of Perspectival Representations and the Role of Gestures in Perceptually Anchoring Inner Speech

Abstract: It has long been observed that gestures do not occur only in communication with others; people also gesture when alone. A common explanation of this phenomenon consists in arguing that verbal thoughts and gestures are better suited to encode and process information, respectively, in propositional and sensorimotor/visuospatial formats. While this is correct, research on inner speech or verbal thought suggests that it provides much more than a format to encode information. Here, I address this issue by suggesting that inner speech and gestures may be concomitantly employed for other reasons than providing complementary ways of processing different kinds of information. More specifically, the focus is on how multimodal self-directed communication involving both gestures and (inner) speech can be used to represent and manipulate different kinds of perspectives: linguistic-attitudinal and sensorimotor. Language in the service of reasoning, taking the form of dialogic inner speech, can be used to recruit alternative and possibly conflicting perspectives on a certain matter, whereas gestures being tightly connected to the activation of sensorimotor mental simulations can be used to express and manipulate iconic and imagistic perspectives. I conclude by suggesting that taking into consideration this form of synergy between gestures and inner speech based on different kinds of perspectives may help to illuminate the role of gestures in perceptually anchoring inner speech.

Keywords: representation, gestures, perspective, inner speech

1 Introduction

Whereas inner speech is increasingly attracting interest across disciplines, it is rarely considered in relation to other forms of self-directed communication. And, while the debate on the functions of inner speech is a lively one, it is taken for

Francesco Fanti Rovetta, Ruhr University Bochum, Bochum,
e-mail: Francesco.FantiRovetta@ruhr-uni-bochum.de

https://doi.org/10.1515/9783110785845-015

granted that whatever inner speech does, it does that alone. At the same time, studies on the use of gestures accompanying inner speech often downplay the many different functions attributed to inner speech, and focus solely on its role in encoding information in a propositional format. However, both these approaches run the risk of ignoring important insights into the complementary use of inner speech and gestures.

In surveying the relation between verbal thoughts and gestures, various studies have focused on the role of gesturing in representing spatial and visual information. It has been argued that verbal thought and gestures complement each other by providing different conceptualization formats (Kita, Alibali, and Chu 2017); thus, better performances can be achieved by—for example—expanding the problem-solving space (Kita 2000; Kita, Alibali, and Chu 2017). However, recently, various studies on inner speech have shown that it does much more than encoding thought-content: inner speech also affords to engage in a wide range of activities, from decision-making to self-reflection, from problem-solving to self-motivating (Morin and Racy 2022 and Martinez-Manrique and Vicente 2015).

A richer understanding of the functions of verbal thoughts could broaden our understanding of the relation between gestures and inner speech and provide novel experimental hypotheses. By considering inner speech in the form of an inner dialogue, this chapter aims to propose a new hypothesis on why we may engage in multimodal self-directed communication, that is, the combined use of inner speech and co-thought gestures, beyond the benefits provided by combining different formats to encode information.

In the following, I will, in a first step, argue that the research on the synergy of inner speech and gestures for cognitive purposes may benefit from a richer understanding of what inner speech is used for; and then in a second step, I will illustrate this idea by claiming that multimodal self-directed communication can facilitate the recruitment of different kinds (i.e., linguistic and perceptual) of perspectival representations.

To this end, I briefly expound on the functions discussed in relation to inner speech, focusing particularly on Neo-Vygotskian approaches. Successively, I explore two views regarding how gestures are generated: the speech production hypothesis (Chu and Kita 2016 and McNeil 2005) and the Gestures as Simulated Action framework (Hostetter and Alibali 2018), and I consider how they account for the combined use of inner speech and gestures. I conclude by arguing that the synergistic use of gestures and inner speech may be useful to represent different kinds of perspectives and I conjecture some of the cognitive benefits that such collaboration may offer.

2 What Do We Do with Inner Speech?

What do we use inner speech for? The question is strictly related to the cognitive functions of language. Indeed, language is not only used for communication but can be employed for cognitive purposes. Empirical results have shown that inner speech plays a role in several cognitive functions: improving short-term memory through inner rehearsal (Baddeley and Hitch 1974), reinforcing executive functions and self-control (Cragg and Nation 2010 and Kompa and Mueller 2020), scaffolding problem-solving (Damianova, Lucas, and Sullivan 2012). Moreover, inner speech may be involved in making thoughts conscious (Carruthers 2015 and Skipper 2022). Carruthers (2015) for example argues that by providing a sensory format that can be targeted by attentional resources, the content of inner speech episodes can enter the working memory system and be consciously accessed.

Inner speech also provides a supplementary set of self-generated stimuli which can override environmental influences. This capacity of inner speech to keep working memory busy, thus overshadowing environmental stimuli, is observable in self-control tasks. During the classic marshmallow experiment (Mischel and Ebbesen 1970), children who succeeded in waiting to eat the treat in order to receive a double quantity later engaged in all kinds of linguistic-based attention-diverting strategies, such as singing and repeating the experiment's instructions. The use of mantras may also be a case in point. In this respect, inner speech has been framed as a form of autostimulation (Dennett 1991 and Clowes 2006), which can be used to direct attention (Clark 2008) and can also prompt us to achieve the desired goal. For example, the linguistic rehearsal of a set of instructional nudges (Sutton 2007) can guide attention, leading to a smoother completion of the task. A former alcoholic at a party may say to themselves, "Whatever happens stick to soft drinks" and when the time comes for toasts, facing the temptation, they may whisper to themselves: "I said no alcohol."

This example shows that the uses of inner speech may not be different from those of social speech. Indeed, according to Neo-Vygotskian approaches, the functions of inner speech are continuous with those of social speech. There is not a single proper function of linguistic thinking for cognitive processes, there are many. The reason for this is to be found in the developmental approach adopted by Neo-Vygotskians and by Vygotsky himself. Inner speech is the result of the process of interiorization of language which begins with social communication. Shortly after having started to engage in linguistic interactions, children begin to use outer speech in a self-directed manner. They overtly talk to themselves as if they were talking to someone else, they give themselves directions, describe their actions, or engage in self-appraisal. Private or egocentric speech has plausibly the same functions as inner speech, with the single downside of being hearable by others. Between the age of 3–5 years old, there is a reduction in private self-talk, supplemented plausi-

bly by inner speaking, although some instances of whispers/muttering are still observable. In this view, inner speech is the result of the completion of the process of internalization of self-directed speech, even though private speech does not disappear during adulthood (Diaz and Berk 1992 and Winsler and Naglieri 2003; for a critical view, see Gregory 2022).

According to this picture, through the process of internalization, children do not simply learn to generate linguistic mental imagery, but to talk to themselves as interlocutors and establish a dialogic relation with themselves. They can give themselves instructions on how to reach some cookies (Vygotsky 1934/2012), ask themselves why they are in a bad mood, motivate themselves and so on. Inner speech is flexible enough to range from the repetition of strings of words, to open-ended quasi-conversations.

Authors adhering to the Vygotskian framework depict the cognitive use of language not as a specialization of some of its functions, but as the internalization of communicative practices which are repurposed to serve various cognitive demands. In the Vygotskian terminology, inner speech is one among many psychological tools that scaffold the process of semiotic mediation, in which socially shared and culturally inherited sign systems (such as verbal language or gestures) are recruited for cognitive purposes.

Tomasello (2019) for example has explicitly adopted a Neo-Vygotskian framework to account for the development of humans' cognitive abilities. In his proposal, a central step in the development of mature cognitive functions is represented by the acquisition and internalization of language. Tomasello claims the process of self-directing language involves role reversal imitation. Children learn to guide their own behaviors by assuming the perspective of sapient instructors and by simulating some directives they would expect to receive from them. In the case of language, they learn to talk to themselves to appropriate the cognitive benefits provided by linguistic interactions. Role reversal imitation scaffolds self-regulation, and language, in the form of private or inner speech, provides a suitable platform for this process to be realized (see Vygotsky 1930/1978).

Tomasello's argument shows that inner speech is not merely a format to encode information, but it also scaffolds important cognitive achievements. Other Neo-Vygotskian works on inner speech have similarly suggested that through inner speech it is possible to engage in simulated linguistic interactions, such as asking questions and then answering them (Clowes 2006), debating, engaging in decision-making processes and so on. Such dialogue-like verbal thinking (also known as dialogic inner speech; Fernyhough 2008) requires the temporary recruitment of alternative and possibly conflicting perspectives ultimately producing less biased and one-sided cognitive results. In writing an academic paper, for example, it is helpful to try as much as possible to anticipate, accommodate and answer eventual

criticisms. Clearly, this process requires to adopting others' viewpoints, background beliefs and assumptions and using them to anticipate potential criticisms.

Similarly, recent works on dialogic inner speech have claimed that it may be used for cognitive functions such as self-knowledge, hypothetical reasoning, and narrative thinking (Morin 2018). In these cases, an exploratory role is attributed to inner dialogue and to its capacity to recruit different perspectives (Fernyhough 2008 and Kompa and Mueller 2022).

However, many scholars who discuss the complementarity of gestures and inner speech (or verbal thought) seem to attribute to the latter a much narrower range of functions, describing it mainly as a means for encoding information (e.g., Kita, Alibali, and Chu 2017). Encoding information in a propositional format is clearly afforded by linguistic thinking, but, at least from a Neo-Vygotskian perspective, it is far from being the only important cognitive function of inner speech.

3 Co-Speech and Co-Thought Gestures

In the scientific literature on gestures, a common distinction is that between co-speech and co-thought gestures. Generally speaking, the former is employed for communicative purposes and to accompany verbal communication, while the latter is used to serve cognitive goals and accompany thoughts.

As for co-speech gestures, the focus is on how they may be used to convey meaning and facilitate conversation. Far from being a mere add-on to verbal communication, gestures may have been all our far ancestors had to communicate at first. The proponents of the gesture-first hypothesis argue that gesturing may be at the origin of the evolution of language. There is indeed some important empirical evidence that shows how easily humans can communicate using only gestures, as shown by the quick emergence and conventionalization of gestures into a sign language observed in modern times, among both deaf and hearing people.[1] Others, who oppose the gesture-first hypothesis, argue for the unity of gestures and speech since their origins (McNeil 2005). The concomitant use of gestures and speech would have to be traced back to the origin of language itself. Whatever the origin of language would prove to be, there is little doubt regarding the complementary of

1 See Kegl, Senghas, and Coppola (1999) on how Nicaraguan Sign Language formed spontaneously among deaf schoolchildren. As for hearing people, secondary sign languages develop when communication is necessary, but speaking is not effective or allowed. Some instances can be found among workers in sawmills, in which noises are too loud to allow vocal exchanges, or in monastic orders which follow the precept of silence (Goldin-Meadow and Brentari 2015).

language and gestures in their capacity for conveying meaning, in that they are: "co-expressive of the same idea yet are opposite semiotically; this is dual semiosis: the gesture is global, synthetic and non-combinatoric; the speech bits are analytic and combinatoric" (McNeil 2017, 80). Thus, gestures and speech are often concomitantly co-opted to convey the same information in different and complementary ways.

Assuming that gestures and speech are co-expressive of the same idea we may then ask whether we need to conceive of it as the content of a pre-existing mental state awaiting to be conveyed and gestures and speech serve as communication channels, or whether they themselves are ways of thinking and creating meaning, while, in social contexts, simultaneously transmitting it. In support of this second interpretation, some researchers, adopting an embodied and extended framework (see Pouw, Nooijer, van Gog, Zwaan, and Paas 2014 for a discussion), have shown that gestures also allow the externalization and subsequent manipulation of spatial representation. Accordingly, instead of mentally imagining the rotation of a cube and using gestures to simulate it, we can depict the rotation with gestures while using the position of the hands to keep track of the rotational phases, thus freeing working memory capacity and offloading the task to the movement of the hands.

If this is the case, at least in some instances, the difference between co-thought gestures and co-speech gestures becomes blurry. Co-speech gestures are also co-thought gestures insofar as communicating becomes an opportunity for thinking through gestures and speech, and co-thought gestures are co-speech gestures insofar as they accompany inner speech. This reading is in line with the Vygotskian framework discussed above: co-thought gestures, just like inner speech, are the result of the re-functionalization of a socially shared and culturally inherited communicative sign system to serve cognitive purposes: gesturing, originally a form of communication, can also be a form of reasoning, rather the expression of something happening in the mind.

Moreover, the Vygotskian framework adopted so far could be extended and radicalized by adopting the suggestion of one of his contemporaries: George Herbert Mead (1934/1972). Mead opposed Wundt's understanding of gestures as the expression of underlying mental states and understood them as functional in constituting the mental state itself. This approach is more radical than Vygotsky's insofar as it assumes communication and sociality as constitutive of cognition and consciousness (Morin 2009, 400, proposes a Meadian approach to inner speech). Whereas for Vygotsky and Neo-Vygotskians socially shared sign systems are tools that transform pre-existing cognitive capacities and allow for communication, according to Streeck, Mead's approach to gestures goes one step further: "Communication is not a product of individual psyches or minds; rather, psyches and minds are epiphenomena of the social process of communication" (2009, 14). However, what matters to present discussion is that for both approaches, these two modes of communication, speech and

gestures, become, by being self-directed, forms of thinking rather than mere modalities for expressing pre-formulated thoughts.

At this point we may summarize the theoretical options described so far regarding theories of gestures into two gestures are either 1) expression of or 2) constitutive of cognitive processes, and, assuming 2), whether the cognitive benefits of combining them with inner speech depend a) only on the combination of different encodings, or b) also on something else. Having argued for 2), the rest of this chapter is devoted to arguing for b), identifying the "something else" as the capacity of multimodal self-directed communication of scaffolding different kinds (perceptual and linguistic) of perspectives. However, to better understand how this is the case, it is first necessary to consider briefly how gestures are generated.

4 How Are Gestures Generated?

Assuming that the distinction between co-speech and co-thought gestures has in some cases a heuristic value (co-thought gestures can be self-directed co-speech gestures), these two kinds of gestures plausibly rely on similar cognitive resources. Indeed, it has been argued that the many similarities between co-speech and co-thought gestures imply that they are generated by common mechanisms. For example, the rate of both co-speech and co-thought gestures is positively correlated to the difficulty of the task—either cognitive or communicative—and it drops when the task is completed. Moreover, people using co-speech gestures more frequently than others also generate co-thought gestures more often (Chu and Kita 2016). These elements suggest a common origin for both kinds of gestures, but there is no consensus on the cognitive processes producing them. Two frameworks are relevant in this regard: theories that embed gesture generation within speech production processes and theories that embed gesture generation within action planning.

According to the speech production hypothesis, gestures and speech have a common origin in the speech production process (McNeil 2005). The idea is that gestures do not partake in communication as extrinsic and accompanying components, rather they are fundamental elements and are generated by the same cognitive mechanisms that originate speech. Whereas in the case of co-speech gestures the speech production process would produce overt vocalization, in the case of co-thought gestures speech remains covert (i.e., inner speech).

The evidence that gestures are sometimes recruited for lexical retrieval (Morsella and Krauss 2004) as well as the evidence for non-verbal semantic processing

in co-speech gestures support the speech production hypothesis. Nonetheless, the speech production hypothesis faces some challenges. For it to be correct, we would expect that overloading the speech production mechanisms by disrupting inner speech would affect the execution of gestures, which does not appear to be the case (Chu and Kita 2016). Moreover, while this hypothesis may explain gestures generated concomitantly with speech, either overt or covert, it may be in an awkward position to explain co-thought gestures when thought is not verbal, such as the use of gestures to facilitate mental rotations of a figure.

An example of the competing paradigm according to which gestures generation is to be found in the cognitive mechanisms supporting action planning is the Gesture as Simulated Action (GSA) framework. According to it, gestures arise from embodied sensorimotor mental simulation (Hostetter and Alibali 2018). By simulations, the two authors refer to "the activation of motor and perceptual systems in the absence of external input" (Hostetter and Alibali 2018, 722). Thus intended, these simulations "are predictive, in that they activate the corresponding sensory experiences that result from particular actions" (Hostetter and Alibali 2018, 722), but it is not specified whether they can be consciously accessed and how they relate to the similar notion of mental imagery. Hostetter and Alibali resort to recent studies on embodied representations to argue that simulations of actions during speaking and thinking "involve motor plans that are the building blocks of gestures" (2018, 722) and thus the "form of gesture presumably depends on the nature of the underlying imagery" (2018, 728).

This framework is grounded on the idea that information can be represented either in a propositional or in an imagistic format. When information is represented verbally or propositionally no action simulation is activated and thus the probability of gesture execution is lower. Conversely, when information is represented in a visuospatial or imagistic format, the motor and perceptual systems are activated and—assuming other criteria are met—some of the features of the simulated action are expressed through gestures.

At first glance, the GSA framework seems to contradict the idea—presented in the section above—that gestures are themselves constitutive of cognitive processes rather than expressing internal mental states. This is not necessarily the case. Indeed, while the authors suggest, based on vast empirical evidence, that gestures are the expression of sensorimotor mental simulation, this does not exclude that gestures in turn can influence mental states, an idea for which there is also considerable empirical evidence (Goldin-Meadow 2016 and Goldin-Meadow, Nusbaum, Kelly, and Wagner 2001).

Of interest for our present discussion is one of the findings discussed in support of the GSA framework. How iconic (or representational) gestures—that is, gestures resembling the denoted object—are executed depends on the perspective

assumed (i.e., the perspective represented in the mental simulation) by the person performing them. The perspective assumed in gesturing an action can be that of the person undertaking the action, but it can also be that of an observer. For example, gesturing of throwing a ball from a first-person perspective (or character-viewpoint) involves simulating the act of throwing, while gesturing the same event from an observer perspective involves tracing the trajectory of the ball with the hands. Unsurprisingly, being perspectival is one key property of iconic gestures (Hassamer and Winter 2018).

5 Different Kinds of Perspectives in Multimodal Inner Speech

So far, I have focused on how self-directed communication, whether through inner speech, gestures or both, can represent different kinds of perspectives or viewpoints. But how do they differ? The hypothesis I intend to suggest here is that gestures represent iconic or imagistic perspectives, while inner speech represents perspectives expressed linguistically. As a first broad characterization of the notion of perspective, we can follow Camp, who describes them as "modes of interpretation: open-ended ways of thinking, feeling, and more generally engaging with the world and certain parts thereof" (Camp 2013, 335–336).

According to the GSA framework, the perspectival nature of iconic gestures is a direct manifestation of the sensorimotor mental simulation from which the gestures are generated. Perspectives expressed by gestures involve sensorimotor information, they represent the perceptual viewpoints from which a scene or the action is imagined; more difficult is to define linguistic perspectives.

At first sight, a clear difference is that linguistic perspectives, as discussed by Neo-Vygotskians, pertain to what one would say, rather than feel, and to what one believes, desires, hopes rather than perceives.[2] This can be clearly seen if we turn to the discussion above: perspectival representation enacted by gesturing rely on the underlying mental simulation, so that one can infer the perceptual perspective taken by observing the iconic gestures produced. Conversely, by engaging in dialogic inner speech, e.g., in the decision-making process, we may assume other perspectives in order to evaluate whether the decision is sound, asking ourselves for example: "What would my parent/partner/friend do?"

2 Studies on cognitive penetrability of perception show that the line is not always easily drawn. I will leave these considerations on the side.

Here, I endorse an attitudinal definition of linguistic perspective: a linguistic perspective consists in the linguistic expression of one or more attitude in relation to one or more proposition (see also Fernyhough 2009, Gregory 2017). Thus, perspectival representations supported by inner speech act as virtual placeholders for sets of attitudes (e.g., beliefs, assumptions, desires, etc.) and they do not necessarily involve perceptual representation.

Once we have clarified the difference between perspective as expressed linguistically and gestural perspectives, the question is: how do they interact? One preliminary suggestion in this regard comes from a recent study on gestures by Kita, Alibali, and Chu:

> Because thinking in terms of action has different properties from propositional or verbal thinking, gesture offers possibilities and perspectives that propositional or verbal thinking cannot, and therefore, gesture affects thinking in particular ways. (Kita, Alibali, and Chu 2017, 246)

While this claim is far from new in the literature on gestures, relevant to our discussion is the remark that through different formats, different kinds of perspectives can be recruited and influence thinking. A further step forward is warranted by studies on the cognitive effects of representing perceptual perspectives either with gestures or with speech. One of these studies investigated whether children at the age of five who produced character-viewpoint gestures became better narrators at the age of six, seven, and eight compared to children who did not produce such gestures, either by producing only observer-viewpoint gestures or no gestures at all (Demir, Levine, and Goldin-Meadow 2015). Other independent variables (syntax comprehension scores, initial narrative structure scores) were excluded as confounding factors. The authors found that "perspective-taking in gesture predicted subsequent narrative structure whereas perspective-taking in speech did not" (Demir, Levine, and Goldin-Meadow 2015, 676), suggesting that even though it may be possible in principle to represent perceptual perspectives in speech, they may not be cognitively efficacious.

While this study does not tell us whether character-viewpoint gestures have a causal role to play in developing more refined narrative structures later on or rather are the effects of some other mental ability which is causally implicated in such development, it shows functional differences between linguistic perspectives and perspectives in gestures. One difference discussed in the study is that compared to gestures, it is more difficult to infer from speech whether the perceptual perspective taken, if any, is that of the character-viewpoint or the observer-viewpoint, which—I suggest—may be due to the more abstract nature of linguistic perspective-taking, based on attitudes rather than perceptual perspective.

Recall that according to the GSA framework, iconic gestures are generated on the basis of the visual mental imagery of the scene depicted, while inner speech does not require to actively visualize what is said. This may suggest that in simply telling a story, the narrator may be agnostic regarding the perceptual representation involved, or, in other terms, speech underdetermines the visuospatial perspective involved in narrating the scene. On the contrary, by producing iconic gestures—which depend on the activation of a mental simulation—the narrator *necessarily* represents a determined perceptual viewpoint.

From this, we may conjecture that, while through the sole use of inner speech it is possible to recruit linguistic or attitudinal perspectives, gesturing may help to anchor what is being said, by providing a perceptual viewpoint. Thus, engaging in iconic gesturing while inner speaking presumably forces the subject to assume a perceptual perspective on what is said. This may be perhaps less useful in epistemic tasks, such as in refining an argument by evaluating various theoretical positions, but it may prove crucial in others, such as in perceptually anchoring a past episode or a narrative and assuming the perspectives of its characters as well as shifting between them, with all the cognitive benefits that may result from it.

6 Combining Perspectives: The Case of Imagined Interactions and the De-Semanticization of Memory

A task for which multimodal self-directed communication (i.e., the synergistic activation of inner speech and gestures) may be fruitfully recruited is that of imagining interactions. Imagined interactions are a form of "social cognition where individuals imagine anticipated or prior communication encounters with others" (Bodie, Honeycutt, and Vickery 2013, 157). Some of the functions attributed to imagined interactions encompass relationship maintenance, conflict management, rehearsing messages, self-knowledge, emotional release, and compensation for lack of actual social interactions.

In many cases, imagined interactions involve the interplay of different linguistic perspectives. In fact, they could be categorized as instances of dialogic inner speech (Morin 2019). Imagining a conversation with a loved one or anticipating how the question-and-answer session of a future talk will unfold both require engaging in linguistic perspective-taking. At the same time, imagined interactions can be multimodal, in that they can include visuospatial imagery and the perceptual

details of the imagined scene. While imagining a past interaction or anticipating a future encounter I can also visually imagine features of the environment, of myself or the other persons. Therefore, producing iconic gestures while engaging in inner speech in these cases may serve to activate a multimodal mental simulation, in which various kinds of perspective, linguistic-attitudinal and visuospatial, interact.

In the case of remembering past interactions, the synergistic recruitment of both linguistic and perceptual perspectives in multimodal self-directed communication may be functional in reversing the process of semanticization of episodic memory. It is argued that, with time, autobiographical memory tends to shift from episodic memory (memory of a lived, subjective experience endowed with perceptual details) to semantic memory (memory of a fact, encoded semantically), the process is also known as "semanticization" of episodic memories (Cermak 1984 and Eustache and Desgranges 2008). This means that, as time passes, memory regarding how something felt becomes knowledge that something happened. The memory of a beautiful sunset on the ocean, the perceptual recollection of the various shades of red, orange and pink, the sounds of the waves and seagulls, etc., becomes knowledge of having experienced a beautiful sunset. Through the process of semanticization of episodic memory, the episode cannot be mentally re-lived anymore, it becomes a remembered fact.

As previously discussed, since iconic gestures depend on the activation of visuospatial mental representation, engaging in iconic gestures when remembering a past event may force the recruitment of a mental simulation of that event. Gesturing while thinking linguistically may help to recreate the moment as experienced first-personally by activating a visuospatial mental representation of the scene remembered which would be left out in speech alone, thus contrasting the process of semanticization of memory.

Clearly, the content of the mental simulation involved in the execution of gestures may not be accurate to how the events unfolded. Whether the combined use of gestures and inner speech to represent a past episode can help to make the memory of it more perceptually vivid, and the degree of reliability of perceptual details added to memories by recruiting gestural perspectives are matters that require further empirical investigation.

7 Conclusion

I have presented a hypothesis on how inner speech and gestures may be concomitantly co-opted in multimodal self-directed communication to represent different kinds of perspectives. I have argued that, following Vygotsky and Neo-Vygotskians in considering how communication can be self-directed to serve cognitive purposes, inner speech and gestures are in themselves forms of thinking, rather than the mere expression of underlying mental states. Moreover, I have argued that previous research on the use of gestures and inner speech focuses on how the combination of the different encodings of thought-content or information, propositional and visuospatial, enhance cognitive performances, and I have suggested an alternative form of synergy between the two based on the different kinds of perspectives, involving propositional attitudes and visuospatial perceptual details, they represent.

The hypothesis presented requires further empirical corroboration in order to better understand whether and how gestures may perceptually anchor inner speech as well as to better understand the cognitive benefits their synergistic use may provide.

References

Baddeley, Alan D. and Graham Hitch. 1974. "Working Memory." In Bower, Gordon H. (Ed.). *Psychology of Learning and Motivation*. Volume VIII, 47–89. Cambridge: Academic Press.

Bodie, Graham D., James M. Honeycutt, and Andrea J. Vickery. 2013. "An Analysis of the Correspondence Between Imagined Interaction Attributes and Functions." *Human Communication Research* 39 (2): 157–183.

Camp, Elisabeth. 2013. "Slurring Perspectives." *Analytic Philosophy* 54 (3): 330–349.

Carruthers, Peter. 2015. *The Centered Mind*. Oxford: Oxford University Press.

Cermak, Laird S. 1984. "The episodic-semantic distinction in amnesia." In Squire, Larry R. and Nelson Butters (Eds.). *Neuropsychology of Memory*, 55–62. New York: Guilford Press.

Chu, Mingyuan and Sotaro Kita. 2016. "Co-thought and co-speech gestures are generated by the same action generation process." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42 (2): 257–270. https://doi.org/10.1037/xlm0000168.

Clark, Andy. 2008. *Supersizing the Mind, Embodiment, Action and Cognitive Extension*. Oxford: Oxford University Press.

Clowes, Robert. 2006. *Beyond situated action: A neo-Vygotskian theory of thinking and language internalisation*. Dissertation. Brighton: University of Sussex.

Cragg, Lucy and Kate Nation. 2010. "Language and the Development of Cognitive Control." *Topics in Cognitive Science* 2: 631–642.

Damianova, Maria, Marilyn Lucas, and Gavin Sullivan. 2012. "Verbal Mediation of Problem Solving in Pre-Primary and Junior Primary School Children." *South African Journal of Psychology* 42: 445–455.

Demir, Özlem E., Susan C. Levine, and Susan Goldin-Meadow. 2015. “A tale of two hands: Children’s early gesture use in narrative predicts later narrative structure in speech.” *Journal of Child Language* 42: 662–681.

Dennett, Daniel. 1991. *Consciousness explained*. New York: Back Bay Books.

Diaz, Rafael M. and Laura E. Berk. 1992. *Private speech: From social interaction to self regulation*. Hillsdale: Erlbaum.

Eustache, Francis and Béatrice Desgranges. 2008. “MNESIS: Towards the integration of current multisystem models of memory.” *Neuropsychological Review* 18 (1): 53–69.

Fernyhough, Charles. 2008. “Getting Vygotskian about theory of mind: Mediation, dialogue, and the development of social understanding.” *Developmental Review* 28 (2): 225–262.

Fernyhough, Charles. 2009. “Dialogic thinking.” In Adam Winsler, Charles Fernyhough, . & Ignacio Montero (Eds.). Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation, New York: Cambridge University Press.

Frankish, Keith. 2018. “Inner Speech and Outer Thought.” In Langland-Hassan, Peter and Agustín Vicente (Eds.). *Inner Speech: New Voices*. Oxford: Oxford University Press.

Goldin-Meadow, Susan. 2016. “Using our hands to change our minds.” *Wiley Interdisciplinary Reviews: Cognitive Science* 8 (1–2): e1368. https://doi.org/10.1002/wcs.1368.

Goldin-Meadow, Susan and Diane Brentari. 2015. “Gesture, sign, and language: The coming of age of sign language and gesture studies.” *Behavioral and Brain Sciences* 40: e46. https://doi.org/10.1017%2FS0140525X15001247.

Goldin-Meadow, Susan, Howard Nusbaum, Spencer D. Kelly, and Susan Wagner. 2001. “Explaining math: Gesture lightens the load.” *Psychological Science* 12: 516–522.

Gregory, Daniel. 2017. Is Inner Speech Dialogic? Journal of Consciousness Studies, Volume 24, Numbers 1–2, 2017, pp. 111–137(27)

Gregory, Daniel. 2022. “How not to decide whether inner speech is speech: Two common mistakes.” *Phenomenology and the Cognitive Sciences* 23: 231–252. https://doi.org/10.1007/s11097-022-09814-w.

Hassamer, Julius and Bodo Winter. 2018. “Decoding Gestural Iconicity.” *Cognitive Science* 42 (8): 3034–3049. https://doi.org/10.1111/cogs.12680.

Hitch, Graham J., Michael E. Woodin, and Sally Baker. 1989. “Visual and phonological components of working memory in children.” *Memory & Cognition* 17: 175–185.

Hostetter, Autumn B. and Martha W. Alibali. 2018. “Gesture as simulated action: Revisiting the framework.” *Psychonomic Bulletin & Review* 26: 721–752.

Kegl, Judy, Ann Senghas, and Marie Coppola. 1999. “Creation through Contact: Sign Language Emergence and Sign Language Change in Nicaragua.” In DeGraff, Michel (Ed.). *Language Creation and Language Change: Creolization, Diachrony, and Development*, 179–237. Cambridge: MIT Press, Bradford Books.

Kendon, Adam. 2014. “Semiotic diversity in utterance production and the concept of ‘language’.” *Philosophical Transactions of the Royal Society B* 369: 20130293. https://doi.org/10.1098/rstb.2013.0293.

Kita, Sotaro. 2000. “How representational gestures help speaking.” In McNeill, David (Ed.). *Language and gesture*, 162–185. Cambridge: Cambridge University Press.

Kita, Sotaro, Martha W. Alibali, and Mingyuan Chu. 2017. “How do gestures influence thinking and speaking? The Gesture-for-Conceptualization hypothesis.” *Psychological Review* 124: 245–266.

Kompa, Nikola and Jutta L. Mueller. 2020. “How Abstract (Non-embodied) Linguistic Representations Augment Cognitive Control.” *Frontiers in Psychology* 11: 1597.

Kompa, Nikola and Jutta L. Mueller. 2022. “Inner speech as a cognitive tool—or what is the point of talking to oneself?” *Philosophical Psychology*. https://doi.org/10.1080/09515089.2022.2112164.

Martínez-Manrique, Fernando and Agustín Vicente. 2015. “The activity view of inner speech.” *Frontiers in Psychology* 6: 232.

McNeill, David. 2005. *Gesture and thought*. Chicago: University of Chicago Press.

McNeill, David. 2017. “Gesture-speech unity, what it is, where it came from.” In Church, Ruth B., Martha W. Alibali, and Spencer D. Kelly (Eds.). *Why Gesture? How the hands function in speaking, thinking and communicating*, 77–102. Amsterdam: John Benjamins.

Mead, George H. 1934, 1972. *Mind, self, and society from the standpoint of a social behaviorist*, edited by Charles W. Morris. Chicago: University of Chicago Press.

Mischel, Walter and Ebbe B. Ebbesen. 1970. “Attention in Delay of Gratification.” *Journal of Personality and Social Psychology* 16 (2): 329–337.

Morin, Alain. 2009. “Inner speech and consciousness.” In Banks, William P. (Ed.). *Encyclopedia of consciousness*, 389–402. Amsterdam: Elsevier.

Morin, Alain. 2018. “The self-reflective functions of inner speech: Thirteen years later.” In Langland-Hassan, Peter and Agustín Vicente (Eds.). *Inner speech: New voices*, 276–298. Oxford: Oxford University Press.

Morin, Alain. 2019. “When Inner Speech and Imagined Interactions Meet.” *Imagination, Cognition and Personality* 39 (4): 374–385. https://doi.org/10.1177/0276236619864276.

Morin, Alain and Famira Racy. 2022. “Frequency, Content, and Functions of Self-Reported Inner Speech in Young Adults: A Synthesis.” In Fossa, Pablo (Ed.). *Inner Speech, Culture & Education*, 147–170. Berlin: Springer.

Morsella, Ezequiel and Robert Krauss. 2004. “The Role of Gestures in Spatial Working Memory and Speech.” *The American Journal of Psychology* 117: 411–424.

Pouw, Wim T. J. L., Jacqueline A. de Nooijer, Tamara van Gog, Rolf A. Zwaan, and Fred Paas. 2014. “Toward a more embedded/extended perspective on the cognitive function of gestures.” *Frontiers in Psychology* 5: Art. 359. https://doi.org/10.3389/fpsyg.2014.00359.

Skipper, Jeremy I. 2022. “A voice without a mouth no more: The neurobiology of language and consciousness.” *Neuroscience & Biobehavioral Reviews* 140: 104772. https://doi.org/10.1016/j.neubiorev.2022.104772.

Streeck, Jürgen. 2009. *Gesturecraft: The manu-facture of meaning*. Amsterdam: John Benjamins.

Sutton, John. 2007. “Batting, Habit and Memory: The Embodied Mind and the Nature of Skill.” *Sport in Society* 10 (5): 763–786.

Tomasello, Michael. 2019. *Becoming Human: A Theory of Ontogeny*. Cambridge: Harvard University Press.

Vygotsky, Lev S. 1930–1935, 1978. *Mind in society: Development of higher psychological processes*. Cambridge: Harvard University Press.

Vygotsky, Lev S. 1934, 2012. *Thought and language*, edited by Alex Kozulin. Translated by Eugenia Hanfmann, Gertrude Vakar, and Alex Kozulin. Cambridge: MIT Press.

Winsler, Adam and Jack Naglieri. 2003. “Overt and covert verbal problem-solving strategies: Developmental trends in use, awareness, and relations with task performance in children aged 5 to 27.” *Child Development* 74: 659–678.