THEORIA ET HISTORIA SCIENTIARUM, VOL. XIX
Ed. Nicolaus Copernicus University 2022

DOI: http://dx.doi.org/10.12775/ths.2022.002

Angelo Delliponti
Nicolaus Copernicus University
angelo.d.delliponti@gmail.com
https://orcid.org/0009-0009-3841-4173

Motor Simulation and Ostensive-inferential communication: insights and clarifications

Abstract. In the article titled "Motor Simulation and Ostensive-Inferential Communication", a theoretical model of how motor simulation is a mechanism that underlies language acquisition is described. It is argued that motor areas might play a role in both the recognition of linguistic communicative and informative intentions in infants, by activating brain regions dedicated to speech processing. In this paper, I will extend the position taken there (i) by connecting my model to the features of infant-caregiver interaction in speech perception, (ii) by explaining the process that causes brains to create networks between speech areas and the motor cortex, and (iii) by showing how the most influential mindreading models can be made compatible with both the embodied simulation theory and with the cognitive abilities in children.

Keywords: language acquisition; ostensive communication; motor simulation; intentions; mindreading.

1. Introduction

In the article titled "Motor Simulation and Ostensive-Inferential Communication" (Delliponti, 2022), an embodied model of ostensive communication (Scott-Phillips, 2014; Sperber & Wilson, 1986) in language acquisition is described. The main goal of the paper was to outline a model of how evidence regarding motor cortex activation during speech listening plays a role in the detection of ostensive cues typically involved in linguistic communication: in a few words, seeking a meeting point between the ostensive model of communication and *motor simulation* (MS), and showing its role in language acquisition. The ostensive-inferential model, also known as *ostensive communication* (OC), explains how people communicate by expressing and recognizing their intent to communicate and inform others about something (Sperber & Wilson, 1986). So, according to this model there is a cognition-based distinction between communicative and informative intentions. With informative intentions, we attempt to make our intended message (its content) clear to our recipient. The information provided to the interlocutor serves as the content of an informative intention and corresponds to the changes that the sender hopes to bring about in the recipient's mental representations. In the case of communicative intentions, we aim to make clear to the intended recipients the very fact that we want to communicate. *Ostension* as an offer of cues and *inference* as an interpretation of the cues are essentially what "ostensive-inferential" means (Scott-Phillips, 2014).

Therefore, it is possible that an embodied mechanism exists for the recognition of linguistic communicative intentions during the daily communicative interactions. However, we know that the recognition of ostensive signals can occur in different ways, not only in non-verbal communication, but also in the verbal one: for example, through the perception of facial expressions or the recognition of gestures (Wilson & Sperber, 2002), or through eye contact (Csibra, 2010). For this reason, the main goal of the paper (Delliponti, 2022) was to propose a model of embodied ostensive communication in relation to language acquisition, thus restricting the scope of application of the model – and of the hypothesis – to language development. I will do the same in this paper as well.

Specifically, my hypothesis was based on an interpretation following the evidence concerning MS, i.e., that the activation of the phono-articulatory areas of the brain following listening to phonemes or, more generally, speech, has a role in the recognition of communicative intentions and that the activation of motor areas that respond to speech content (e.g., an action word), namely the somatotopic activation of the brain area related to a specific action (e.g., the primary motor area involved in leg movement after hearing "to kick"), is involved in the recognition of informative intentions. I named the two processes, respectively, "phono-articulatory simulation" and "semantic simulation"; in turn, these mechanisms have an important role in language acquisition. Building on that work (Delliponti, 2022), here I will: (i) outline a model of the role of phono-articulatory simulation in baby talk and explain how this role is important for the recognition of ostensive cues in infants and for language acquisition; (ii) explain and outline in detail the role of semantic simulation in the recognition of informative intentions, how it is the result of associative learning and what is its role in the acquisition of action words; (iii) suggest which mindreading models fit best, after introducing the main ones, in order to describe how MS can be involved in understanding communicative and informative intentions.

2. Motor cognition and intentions

In this section, I will show some of the evidence regarding the role of motor cognition in the recognition of intentions. This is because there are already theories – in the literature – regarding the role of motor activation in deducing intentions. In the earlier work (Delliponti, 2022), a hypothesis about the role that the activation of motor cortex may have in language learning in infants was proposed. MS is an activation of sensorimotor patterns. In particular, they are *re*-activated regardless of their motor functions and used in cognitive processes unrelated to those for which they evolved (Borghi & Caruana, 2016). The idea behind this mechanism is that *mirror neurons* (MN) enable MS, which is typically viewed from the standpoint of embodied cognition as an automatic system: one hypothesis is that MN, which are located in the premotor cortex, facilitate the activation of the primary motor cortex, and that this is a consequence of a cortico-cortical effect induced by action observation (Fadiga et al., 2005). In fact, there is evidence in macaques that MN fire both when monkeys make goal-directed hand motions and when they observe other humans doing comparable movements (Di Pellegrino et al., 1992): the same mechanism is thought to be activated in humans' ventral premotor cortex, in the homolog region of the F5 monkey area (Fadiga et al., 2005). One of the hypotheses behind the functioning of MN is that they are involved in recognizing others' intentions (Gallese, 2007): however, this idea has been repeatedly criticized over the past ten to fifteen years (Cook et al., 2014; Hickok, 2009). The main misunderstanding on MN, it appears, is related to theories explaining understanding intentions by a homuncular-like functioning (Mikulan et al., 2014), as is the case, for instance, with the hypothesis of direct correspondence, which claims that an action is understood when its observation causes a resonance in the observer's motor system (Rizzolatti et al., 2001); in this instance, the understanding attributed to the mirror system is considered automatic and mandatory (Csibra, 2007). Therefore, it is plausible that MN by themselves are not enough to explain how other people's intentions, or the mental states that underlie the acts they watch, are encoded.

The idea that mirror neurons are involved in recognizing the arrangement of body parts when we see an action, however, is supported by several studies (Thompson et al., 2019a). Additionally, according to Thompson et al. (2019b), the information encoded by mirror neurons is subsequently exploited by multiple brain areas "in order to identify the mental state underlying an observed action" (p. 110). The most recent theories about how the MN work view them as a network that extends beyond the motor cortex and includes other regions of the brain, like those involved in highly complex cognitive functions as mentalization (Salo et al., 2019) The process of deducing the intentions behind an action would therefore involve a combination of bottom-up and top-down mechanisms.

3. Motor simulation and language

Moreover, research indicates how hearing phonemes, words and sentences activates specific motor areas. A TMS experiment (Fadiga et al., 2002) showed that hearing phonemes causes an increase in the motor evoked potentials (MEPs) amplitude recorded from tongue muscles normally involved in producing them. The result was interpreted as an acoustically connected resonance mechanism. This phenomenon was confirmed in a series of studies (Gallese, 2007). In an electromyography experiment by McGuigan and Dollins (1989), it was found that tongue and lip muscles are activated in the same manner during both the production of regular speech and covert speech. In Delliponti (2022), it was proposed that this evidence concerning motor activation at the phono-articulatory level while hearing phonemes, words, etc., can be considered as supporting the hypothesis of a phono-articulatory simulation (see also Fischer & Zwaan, 2008).

Secondly, other studies show evidence for a motor cortex activation sensitive to the content of words. In Martin et al. (1996), it was shown that the left middle temporal gyrus, which is activated during action tasks, as well as the left premotor cortex, which is typically activated when people imagine themselves holding objects in their dominant hand, are both differentially activated when pronouncing tool names as opposed to animal names. Other research demonstrates that exposure to words that denote instruments or actions causes a motor resonance (having the activation of motor areas as an effect). According to research by Hauk et al. (2004), action words that describe movements of the face, arms, or legs, somatotopically activate the fronto-central cortex, supporting the idea that the sensorimotor cortex processes certain aspects of the meaning of action-related words (Pulvermuller, 2005). Similarly to the case of the phono-articulatory effect, it was suggested (Delliponti, 2022) that the evidence concerning somatotopic motor activation, when motor cortex responds to the content of the words, can be considered as supporting the hypothesis of a semantic simulation (Fischer & Zwaan, 2008).

However, my main hypothesis was that phono-articulatory simulation and semantic simulation are mechanisms associated with OC. Specifically, that these processes result from the neural exploitation hypothesis (Gallese, 2003; Gallese & Lakoff, 2005), from which the MS theory originates, and that they deal with the recognition of ostensive signals relating to a specific means of communication, i.e., language: the phono-articulatory simulation as having a role in the recognition of communicative intentions, and the semantic one as having a role in the recognition of informative intentions (both in language).

4. Ostensive signals in infants

The point of my thesis, however, is to explain how motor simulation (phono-articulatory and semantic) plays a role in language acquisition, and what are the details related to the mechanisms involved, specifically, in the recognition of communicative and informative intentions. According to Csibra (2010), infants easily recognize the meaning of ostensive signals that are encoded as communicative intentions. Rather than being the result of the growth of communication abilities, recognizing ostensive signals – in the case of communicative intentions, observing their presence and not necessarily accessing their content – is one of the sources. So, communication development is made possible by the fact that the ability to understand them is innate. Ostensive signals must satisfy the following requirements: clearly identify the infant as the recipient of a communicative act; be discernible to neonates; and elicit a preference for the source. At least three different types of stimuli meet these requirements: direct gaze resulting in eye contact; the specific intonation pattern known as baby talk, motherese or infant-directed speech, that is employed with infants; and contingent reaction to the infant's behavior in a turn-taking way. I claim that this facilitation to recognize ostensive signals in infants might happen also in the case of informative intentions. For the purposes of the next section, I will focus on baby talk.

5. Phono-articulatory simulation in language acquisition

I will present here some of the evidences of how baby talk can play a role in the recognition of communicative intentions and what is the role of motor simulation. There is a specific aspect of how baby talk might be involved in phono-articulatory simulation and, accordingly, in language acquisition. The human hearing system has got special features that enable it to distinguish human voice from background noise (Csibra, 2010). With a bias toward speech, newborns can distinguish between speech and non-speech stimuli. Specialized brain regions support this differentiation, and people are naturally more sensitive than other animals to this form of communication (Vatakis et al., 2008). But hearing speech does not definitely provide the conclusion of being addressed, and differently from eye contact speech does not directly indicate the addressee of communication. You will know the addresser is speaking to you, for instance, if they use your name, welcome you politely, refer to events that are pertinent to your specific situation or to anything you said or did before, and so on. By the way, the issue here is that preverbal infants are unable to decode the message of what one says, while those methods work only in case one can decode the content of a speech. Even though infants are not the ones being spoken to (most of the time), they can hear speech, and given specific cues by the speakers – indicators that make it clear when speakers are speaking to a young child, but not necessarily eye contact – infants are able to recognize that speech is addressed to them.

Moreover, when speaking to preverbal newborns, adults automatically change their prosody (Csibra, 2010). Infant-directed speech, or baby talk, differs from adult-directed speech in pitch, amplitude fluctuation, and speed. Although there are cultural variances, these features of baby talk are universal (Fernald, 1995). It has been suggested that this specific style of speech directed at infants has a number of purposes, including capturing the infants' attention, regulating affect, maybe being a cause of language learning, or simply being a result of talking to infants in emotionally charged situations (Csibra, 2010). So, according to Csibra, "the immediate function of the infant-directed intonation pattern is […] it makes it manifest that the speech is infant-directed. […] the special prosody associated with motherese indicates to the baby that he is the one to whom the given utterance is addressed, and so it serves as an ostensive signal" (ibid., p. 148). It is also likely that this feature, i.e., the preference for baby talk, is innate in humans. So, baby talk "is very effective in orienting infants to the speaker, and mothers use it to achieve exactly this effect" (Csibra, 2010, pp. 148–149). When infants cannot determine that they are being spoken to, based on the speech content, adults often utilize baby talk, which complements infants' sensitivity to it. Basically, this means that baby talk is important for infants in order to acquire language, not necessarily because the features of baby talk help them to understand words, but mainly as infant-directed speech is crucial for them in order to recognize linguistic communicative intentions: in turn, as a side effect, this helps them with language acquisition.

On that note, how is phono-articulatory simulation involved in infants' sensitivity to baby talk? My hypothesis is that the communicative resonance mechanism is crucial to language learning (Delliponti, 2022) because it makes sure that the infant's focus is solely on language and not on other "communication systems". Therefore, the identification of linguistic communicative intentions would involve MS, namely the phono-articulatory one, that is involved in the recognition of communicative intentions. From this point of view, baby talk is a mechanism that facilitates the activation of the phono-articulatory system: as a consequence, when adults resort to baby talk, a greater activation of speech related motor areas should be observed in infants. The content of motor processing (low-level) would then be sent to the mentalizing system, so that the process of recognition of the communicative intention (high-level) would be successful (Salo et al., 2019). In short, the act of communicating is processed by means of the phono-articulatory resonance (Fischer & Zwaan, 2008), and in the case of baby talk, this results in a greater activation of speech related motor areas in infants. My conclusion is therefore that, as a side effect, phono-articulatory simulation could play a very important role in language acquisition.

6. Semantic simulation in language acquisition

With regard to semantic simulation, in my previous paper (Delliponti, 2022), an involvement of motor resonance in the recognition of linguistic informative intentions was suggested. As previously mentioned, this hypothesis is based on the evidence concerning somatotopic activation of the motor cortex responding to the content of the words (Hauk et al., 2004; Martin et al., 1996; Preissl et al., 1995; Pulvermuller et al., 1999), and more specifically, action words or action verbs (Pulvermuller et al., 2005). As claimed by Fischer & Zwaan (2008, p. 837): "*referential motor resonance* occurs when the motor system responds to the content of the communication". The same authors make clear the distinction between phono-articulatory and semantic simulation:

> If a listener's speech motor system responds to hearing the word "kick", then this would be an example of communicative motor resonance; the motor system is simulating the production of the utterance. However, if the leg area of the premotor cortex responds, this would indicate referential motor resonance; the motor system is simulating the action that is being described by the utterance rather than the production of the utterance itself (Fischer & Zwaan, 2008, p. 837).

However, it is necessary to clarify in which sense, and what it means that semantic simulation has a role in the recognition of linguistic informative intentions. Here, one might think that this mechanism is similar or specular to that of phono-articulatory simulation, but on closer inspection, it is possible to see that it is a different process, with different features. It was also claimed (Delliponti, 2022) that semantic simulation is consistent with the notion that our ancestors' environment caused selection pressures in favor of vocal information with action content, as communication and language originated for action (Borghi & Caruana, 2016). This indeed seems consistent with an embodied approach to the origin of language, *embodied* eventually in a weaker and not necessarily in a strong sense.

So, what does it mean that semantic simulation is involved in recognizing informative intentions? We know that associative learning is the mechanism that leads to the sensorimotor processing of verbs, in adults (Cooper et al., 2013; Heyes, 2010), and a similar process happens in infants (7 to 9 months olds) with regard to the processing of action related sounds (Gerson et al., 2015; Paulus et al., 2012, 2013). Moreover, motor areas are activated when action verbs are heard during the early stages of language acquisition (Antognini & Daum, 2019). This means that the processing of action related verbs involves the sensorimotor system in infants. Fargier et al. (2012, p. 889) explain how somatotopic activation of motor areas during the hearing of action words, and mostly verbs, is a consequence of associative learning:

> Since "action words" (mostly verbs) are often acquired and experienced in the context of execution of the depicted actions [...], and given Hebb's postulate that synchronous activity of neurons leads to the formation of neuronal assemblies [...], Pulvermuller suggested that neural networks including perisylvian language areas and motor areas emerge with experience. By means of these shared circuits, perceiving an action word will then automatically trigger activity in motor regions of the brain [...].

Given the associative learning process, a hypothesis is that at an early age the motor system, in conjunction with the mentalizing system, helps to recognize the intention behind an action. It is the theory that combines evidence about MS as a mechanism that helps to provide information about intention (Gallese, 2007), plus the evidence about the role of high-level systems, namely the network consisting of the motor cortex and the brain areas of mentalization (Salo et al., 2019). This leads to recognizing the intention of an action, as well as the action itself.

Consequently, assuming a knowledge of the action already possessed (but not strictly necessary), my thesis about the role of associative learning is that it is possible to acquire a new (action) word by relying on the information contained in the recognition of the intention. As said earlier, this happens because action words are frequently learned in the context of performing the actions shown. Thus, an association is formed between the intention behind an action and the intention behind the word (e.g., to grasp). My hypothesis is that the recognition of the informative intention behind the association of word and action (by the recognition of the intention of the action) helps to consolidate the sense of the word. As a result of the associative learning, there is a somatotopic activation of the motor cortex upon hearing the learned action word. This mechanism is involved in language acquisition and probably plays an important role, considering that infants learn words in stages, with more abstract words coming later, whereas the first verbs they acquire are largely verbs describing observable actions (Antognini & Daum, 2019; Ponari et al., 2018; Reggin et al., 2021). So, semantic simulation is a result of associative learning, that is the mechanism properly at work during the recognition of the informative intention of the action, and the association of the correspondent linguistic informative content (see Figure 1). In my model, it is the associative learning – via recognition of the informative intention – that facilitates the acquisition of action verbs, while semantic simulation (which takes place after the process has occurred) is only a result of learning. Since at the time there is no definitive evidence on the role of semantic simulation, it is not entirely out of place to define it as a "secondary effect".

7. Motor simulation and inference: what kind of mindreading?

In the previous sections I suggested a model of MS and how it plays a role in OC, illustrating the way in which this model plays, in turn, an important role in language acquisition. I will now try to suggest what kind of mindreading might be at work in these specific cognitive processes related to the developmental phase, an issue involved in the broader problem of mindreading in infancy (Butterfill & Apperly, 2013; Carruthers, 2013, 2016; Goldman, 2006; Goldman & Jordan, 2013; Rakoczy, 2012). There are some basic questions relevant to the topics presented here, e.g.: Do newborns have a theory of mind? And if so, what type? Is it explainable within the framework of the "classical" theory of mind, or is it of a different kind? These are clearly nontrivial questions to which, however, attempts have been made in recent years to give some answers; and it will be the experimental work, possibly, to offer new evidence in order to account for the less clear aspects of the theory. However, what I will do in this section is to present some mindreading models and suggest which of them have features compatible with the cognitive resources of early childhood and with the MS model presented here.

To put it simply, there are two main models that describe, in different ways, mechanisms and features of mindreading: the theory-theory (TT) and the simulation theory (ST) (Goldman & Jordan, 2013). Each of these main strands can be divided into two categories characterizing specific modules, distinct or constituting one another's subset, each with certain properties. TT can be divided into *full-blown* theory of mind (FB-ToM) and *minimal* theory of mind (M-ToM) (Butterfill & Apperly, 2013), while ST can be divided into *high-level* simulational mindreading (HL-SM) and *low-level* simulational mindreading (LL-SM) (Goldman & Jordan, 2013). What characterizes the difference between the distinct types of TT (FB-ToM and M-ToM) and ST (HL-SM and LL-SM) is the specific degree of complexity involved, complexity related to the cognitive resources and the processing difficulty implicated in mindreading. Consequently, it is possible that – under certain conditions – each subdivision is addressed to a specific object.

Generally speaking, the theory of mind (ToM) is the ability to infer from others' thoughts, beliefs, and emotions, what their intended action would be, in order to predict it (Byom & Mutlu, 2013). As for the TT, FB-ToM involves the mental representation of propositional attitudes such as beliefs, desires and intentions, e.g.: subjects represent the belief of another agent, such as an object is behind a wall, by holding a second-order belief, namely a representation, and not by adopting or imitating the first-order belief that the object is behind the wall (Lurz et al., 2022). This is a representation about a representation, or *metarepresentation* (see Figure 2). Otherwise, in the case of M-ToM, one of the proposed explanations is that subjects use proxies in order to attribute to agents perceptual states, beliefs or intentions: these proxies are defined by Butterfill & Apperly (2013) as *encountering* and *registration*. Under a limited range of commonplace situations, agents sense an item only when they come into contact with it, and they believe that an object has a certain property only when they register it as having that property (see Figure 3). So, according to the authors (ibid.), encountering and registration are ways to attribute mental states to others without involving any representation about representations; it is enough to process goal-directed actions by representing their outcomes as functions of motions made by a body (and not representing mental states). Hence, in order to possess a M-ToM it is enough to understand bodily movements as "units which are directed to goals" (ibid., p. 614).

As for the ST, it requires first-order beliefs with similar content to the first-order beliefs encoding other agents' actual representations. HL-SM hypothesizes that mindreaders use their own minds to create mental models of their intended targets. When a subject places her cognitive processes in the same "starting-state" as the agent's and, as a result, those processes direct her, this simulation may allow her to predict what the agent will do (Goldman & Jordan, 2013). Importantly, it is mostly a product of imagination and involves a decision-making mechanism (see Figure 4). HL-SM differs from LL-SM as this one, unlike HL-SM, is an automatic process that does not require the use of imagination or a decision-making mechanism (see Figure 5).

Conditions such as the mirroring of disgust and pain, or motor simulation, are automatic processes that directly trigger a reaction in the mindreader / simulator, similar in the content to the state of the agent; they are therefore implicit, low-level representations.

Thus, what kind of mindreading may infants have, compatible with the MS theory presented here, specifically the MS involved in language acquisition? The literature on mindreading has repeatedly underlined how problematic it is to attribute a FB-ToM to newborns, on the basis of the evidence concerning childhood skills on attributing intentions to others (Carruthers, 2013; Rakoczy, 2012); similar issues have also affected the debate on mindreading in non-human animals (Bermúdez, 2009; Lurz et al., 2022). However, based on some groundbreaking studies (Onishi & Baillargeon, 2005; Southgate et al., 2007), we know that pre-verbal infants possess the ability to recognize goals, perceptions, and beliefs, based on some form of sensitivity to false belief tasks. On the basis of what I claimed previously, it would seem reasonable to suppose that the type of mindreading taking place during the phono-articulatory simulation and during the semantic simulation, in infants, is linked to the ST: this also seems obvious given that MS, which is a form of embodied simulation, is based precisely on the ST (Goldman & de Vignemont, 2009). And it also seems reasonable to suppose, on the evidence presented in this paper, that some of the mentalizing tasks can be described with reference almost exclusively to empathic mirroring, i.e., LL-SM. In fact, MS is in all respects a type of LL-SM: the activation of motor areas specialized in the phono-articulatory movements or in the movements of other parts of the body (arms, legs, etc.), as happens during the phono-articulatory simulation and the semantic one (activation that in such cases, as mentioned, is consequent to listening to words or phrases, in one case responding to the communicative act, in the other to the content. Activation which, however, is subsequently inhibited, see Borghi & Caruana, 2016), is an automatic mechanism that does not require the use of imagination or of a decision-making process. What happens is that the motor cortex automatically activates in response to exposure to verbal stimuli, low-level activation which is a type of embodied simulation.

However, what kind of mindreading should we refer to in order to explain the recognition of ostensive cues in infants? What I want to suggest in this final part of the paper is that a simulation approach (both low- or high-level) can be accompanied in several cases by a ToM-based approach, depending on the evidence we have on circuit sharing and activation of different areas of the brain during mentalizing tasks (Lombardo et al., 2010). The recognition of communicative intentions of the type described here in infants (baby talk) occurs through phono-articulatory simulation, which is a type of LL-SM. It may be that this is an entirely implicit mechanism, not requiring any kind of high-level representation. However, the semantic simulation, which takes place through a somatotopic activation of the motor cortex, is a type of SM and consequently LL-SM, but it is possible that the described mechanism of attribution of informative intentions could be accompanied by an activation of brain areas involved in higher-level processing. This is because the process of associative learning during the observation of actions accompanying the learning of related action verbs occurs parallel to a mechanism involving the attribution of goals to the action; this process might need a M-ToM, considering that pre-verbal infants may lack the metarepresentative skills of older children (Butterfill & Apperly, 2013). In fact, as said previously, understanding body actions as units that are directed toward goals is all that we need to possess a M-ToM. However, the same process could also be explained by a HL-SM, which would require a first-order representation, in this case through imagination and a decision-making mechanism. It is therefore likely that the associative learning process underway during the acquisition of action verbs, is initially linked to a MS mechanism that is activated following the observation of the action to which the verb corresponds, an association that would create new connections between linguistic and motor areas. The first part of the process could therefore be exclusively explained with the LL-SM. However, as mentioned, attributing an intention to the observed action could be something that implies the activation of other areas, specialized in mentalization tasks (Lombardo et al., 2010). At a later time, the data processed by low-level areas could therefore be sent to other brain areas dedicated to a higher-level processing. Based on the evidence concerning mindreading in childhood (Butterfill & Apperly, 2013), it may be excluded that infants, up to a certain age, are equipped with a FB-ToM, while this second part of the process is likely to rely on a M-ToM or a HL-SM. The result of this learning could equally exploit the same modules (LL-SM and M-ToM, or LL-SM and HL-SM), and therefore the understanding of an action verb would be a process that implies both an activation of motor areas and of areas more specialized in mentalization tasks. However, although I claim here that semantic simulation (here understood therefore as the outcome of the learning process) has an important role in language acquisition in childhood, only future studies could shed light on the role that the part of semantic simulation relating to low-level activation may have during everyday understanding of action verbs.

To conclude, in order to explain the cognitive processes taking place during the MS needed by infants to recognize the ostensive cues useful for language learning, the best method is not to exclude a type of explanation involving a "mixed" approach, with low- and high-level representations, whether this can be explained entirely through the ST, or whether this process can be explained through mechanisms involving representations of different types, as diverse as those at work in distinct models, as in the case of LL-SM and the M-ToM.

Conclusions

In this paper I tried to clarify the main assumptions advanced in Delliponti (2022), extending their implications, and developing some of the points that had not been sufficiently explored. First of all, I defined what the ostensive-inferential model of communication is, explained the theory behind motor simulation. I then introduced some evidence supporting the theories regarding the role of mirror neurons and motor areas in intention recognition, and the evidence for the role of motor areas in language processing. I suggested that motor activation during words listening and, more generally, utterances, could have a similar role to that of intention recognition during the observation of actions, after having clarified in which sense motor areas are involved in the recognition of intentions, and how these are part of a larger network which also includes areas of mentalization. I then introduced two concepts: phono-articulatory simulation (or communicative motor resonance), which occurs when the speech motor system responds to listening to words, simulating the production of the utterance; and semantic simulation (or referential motor resonance), which occurs when there is a somatotopic activation of motor areas responding to the action content of words, simulating it. I then explained how phono-articulatory simulation plays a role in language acquisition, especially in the case of baby talk, which serves infants as ostensive signals for the recognition of communicative intentions. I then explained how semantic simulation is the result of an associative learning process, also crucial for learning action words (especially verbs), since the learning context, in which the word is presented at the same moment in which the action to which it refers is shown, has a role in the understanding of the informative intention behind the word: motor simulation is involved in recognizing the intention behind the action, that intention is then moved to the word, resulting in a Hebbian learning. After the learning phase, listening to the word will be sufficient to activate the same motor areas involved in the action.

Finally, I presented some mindreading models, all attributable to the theory-theory and simulation theory distinction, suggesting that the simulation processes presented here can be supported in some cases by low-level simulational mindreading alone, in the case of phono-articulatory simulation, or by a mix of low-level and high-level mindreading, in the case of semantic simulation, e.g., low-level plus high-level simulational mindreading, or low-level simulation plus minimal theory of mind.

Conflict of interest
The author reports there are no competing interests to declare.

REFERENCES

Antognini K., & Daum M. M. (2019). Toddlers show sensorimotor activity during auditory verb processing. *Neuropsychologia*, 126, 82–91.

Bermúdez, J. L. (2009). Mindreading in the animal kingdom. *The Philosophy of Animal Minds*, 145–164.

Borghi A., & Caruana F. (2016). *Il cervello in azione*. Bologna: Il Mulino.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28(5), 606–637.

Byom, L. J., & Mutlu, B. (2013). Theory of mind: Mechanisms, methods, and new directions. *Frontiers in Human Neuroscience*, 7, 413.

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language*, 28(2), 141–172.

Carruthers, P. (2016). Two systems for mindreading? *Review of Philosophy and Psychology*, 7(1), 141–162.

Cook R., Bird G., Catmur C., Press C., & Heyes C. (2014). Mirror neurons: from origin to function. *Behavioral and Brain Sciences*, 37, 177–192.

Cooper, R. P., Cook, R., Dickinson, A., & Heyes, C. M. (2013). Associative (not Hebbian) learning and the mirror neuron system. *Neuroscience Letters*, 540, 28–36.

Csibra G. (2007). Action mirroring and action understanding: an alternative account. In: P. Haggard, Y. Rosetti, & M. Kawato, *Sensorimotor Foundations of Higher Cognition. Attention and Performance XII*, 453–459. Oxford: Oxford University Press.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, 25(2), 141–168.

Delliponti, A. D. (2022). Motor Simulation and Ostensive-Inferential Communication. *AVANT. Pismo Awangardy Filozoficzno-Naukowej*, 1, 1–20.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91(1), 176–180.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15(2), 399–402.

Fadiga, L., Craighero, L., & Olivier, E. (2005). Human motor cortex excitability during the perception of others' action. *Current Opinion in Neurobiology*, 15(2), 213–218.

Fargier, R., Paulignan, Y., Boulenger, V., Monaghan, P., Reboul, A., & Nazir, T. A. (2012). Learn-ing to associate novel words with motor actions: Language-induced motor activity following short training. *Cortex*, 48(7), 888–899.

Fernald, A. (1995). Human maternal vocalizations to infants as Biologically Relevant Signals: An Evolutionary Perspective. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, 391.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, 61(6), 825–850.

Gallese, V. (2003). A neuroscientific grasp of concepts: from control to representation. *Philosophical Transactions of the Royal Society B*, 358, 1231–1240.

Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B*, 362(1480), 659–669.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in reason and language. *Cognitive Neuropsychology*, 22, 455–479.

Goldman, A. (2006). *Simulating Minds*. New York: Oxford University Press.

Goldman, A. I., & Jordan, L. C. (2013). Mindreading by simulation: The roles of imagination and mirroring. *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*, 448–466.

Goldman, A., & de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences*, 13(4), 154–159.

Hauk, O., Johnsrude, I., & Pulvermueller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301–307.

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews*, 34(4), 575–583.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, 21, 1229–1243.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., MRC AIMS Consortium, & Baron-Cohen, S. (2010). Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience*, 22(7), 1623–1635.

Lurz, R. W., Krachun, C., Mareno, M. C., & Hopkins, W. D. (2022). Do chimpanzees predict others' behavior by simulating their beliefs. *Animal Behavior and Cognition*, 9(2), 153–175.

Martin, A., Wiggs, C. L., Ungerleider, L. G., Haxby, J. V. (1996). Neural correlates of category-spe-cific knowledge. *Nature*, 379, 649–652.

McGuigan, F. J., & Dollins, A. B. (1989). Patterns of covert speech behavior and phonetic coding. *The Pavlovian Journal of Biological Science*, 24(1), 19–26.

Mikulan, E. P., Reynaldo L., & Ibanez A. (2014). Homuncular mirrors: misunderstanding causality in embodied cognition. *Frontiers in Human Neuroscience*, 8, 1–4.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255–258.

Paulus, M., Hunnius, S., Van Elk, M., & Bekkering, H. (2012). How learning to shake a rattle affects 8-month-old infants' perception of the rattle's sound: electrophysiological evidence for action-effect binding in infancy. *Developmental Cognitive Neuroscience*, 2(1), 90–96.

Paulus, M., Hunnius, S., & Bekkering, H. (2013). Neurocognitive mechanisms underlying social learning in infancy: infants' neural processing of the effects of others' actions. *Social Cognitive and Affective Neuroscience*, 8(7), 774–779.

Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, 21(2), e12549.

Preissl, H., Pulvermuller, F., Lutzenberger, W., & Birbaumer, N. (1995). Evoked potentials distin-guish between nouns and verbs. *Neuroscience Letters*, 197, 81–83.

Pulvermuller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6, 576–582.

Pulvermuller, F., Lutzenberger, W., & Preissl, H. (1999). Nouns and verbs in the intact brain: Evidence from event-related potentials and high-frequency cortical responses. *Cerebral Cortex*, 9, 498–508.

Pulvermuller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, 17, 884–892.

Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology*, 30(1), 59–74.

Reggin, L. D., Muraki, E. J., & Pexman, P. M. (2021). Development of abstract word knowledge. *Frontiers in Psychology*, 12, 686478.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the un-derstanding and imitation of action. *Nature Reviews Neuroscience*, 2, 661–670.

Salo, V. C., Ferrari, P. F., & Fox, N. A. (2019). The role of the motor system in action understand-ing and communication: Evidence from human infants and non-human primates. *Developmental Psychobiology*, 61(3), 390–401.

Scott-Phillips, T. (2014). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Bloomsbury Publishing.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18(7), 587–592.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Thompson, E. L., Bird, G., & Catmur, C. (2019a). Mirror neurons, action understanding and social interaction: implications for educational neuroscience. *Conference Abstract: 4th International Conference on Educational Neuroscience*.

Thompson, E. L., Bird, G., & Catmur, C. (2019b). Conceptualizing and testing action under-standing. *Neuroscience & Biobehavioral Reviews*, 105, 106–114.

Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *Journal of Vision*, 8(9), 14.

Wilson, D., & Sperber, D. (2002). Relevance theory. In: G. Ward, & L. Horn, *Handbook of Pragmatics*, Blackwell.

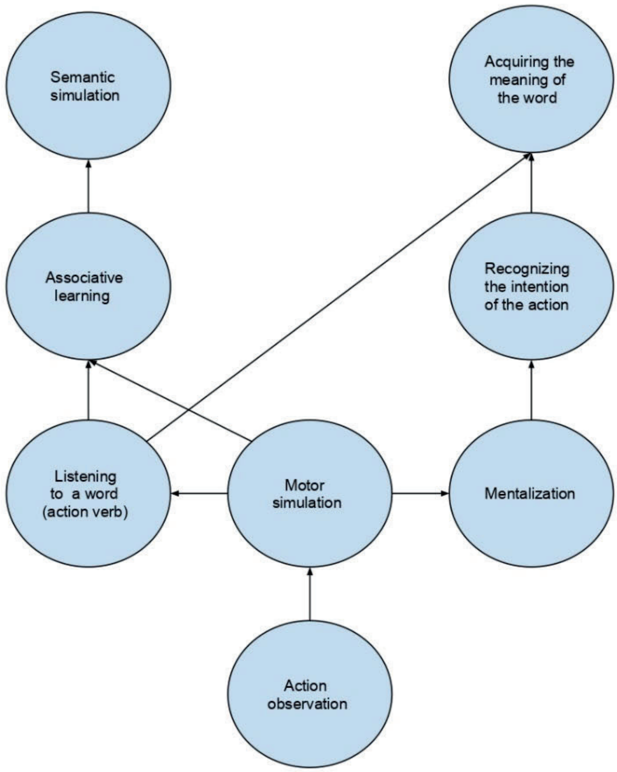

Figure 1. Semantic simulation as a byproduct of recognizing the informative intention in language acquisition

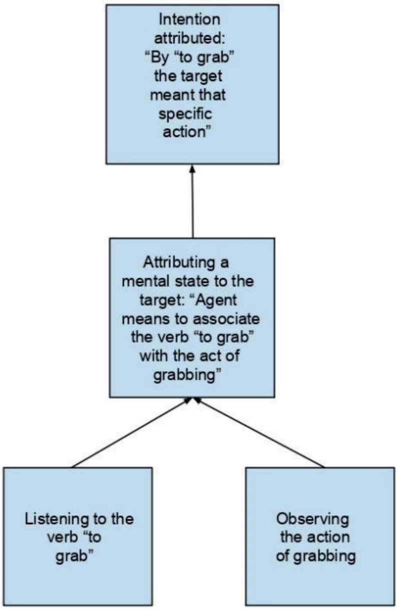

Figure 2. A representation of the full-blown theory of mind

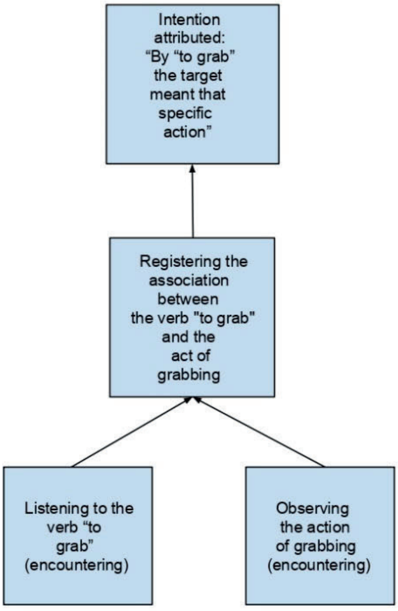

Figure 3. A representation of the minimal theory of mind

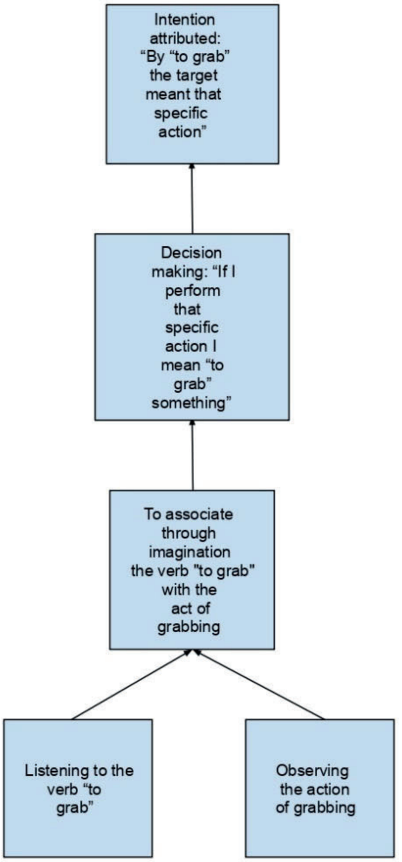

Figure 4. A representation of the high-level simulational mindreading

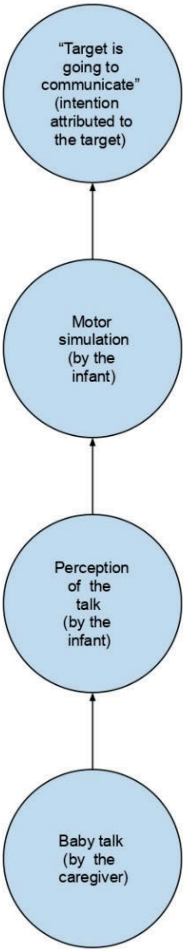

Figure 5. A representation of the low-level simulational mindreading