UNIWERSYTET
MIKOŁAJA KOPERNIKA
W TORUNIU

Angelo D. Delliponti

THE ROLE OF PRAGMATIC COMPETENCE AND OSTENSIVE COMMUNICATION IN LANGUAGE

Doctoral dissertation

Supervisor:

dr hab. Sławomir Wacewicz, prof. UMK

Co-supervisor:

Professor Francesco Ferretti

Toruń 2025

To Professor Dr. Hab. Sławomir Wacewicz
I would like to express my heartfelt thanks for your work, support,
expertise, kindness and invaluable help
in completing this doctoral dissertation.

I would like to thank Professor Francesco Ferretti
for his support and help given to me since the beginning,
without whom I would never have completed my thesis project
and I would never have started this experience.

I would like to thank the employees, PhD students and members
of the Centre for Language Evolution Studies and of the Cosmic Lab
for their work and fruitful cooperation,
which enabled me to prepare my doctoral dissertation.

Thanks to all the co-authors
of the papers included in the thesis,
to which they contributed with essential work for its completion,
and without whom it would not have been the same.

To my fiancée Asia,
Parents, Sister and Friends
I would like to thank you very much for your support
in the creation of this work.

Table of contents

1. **Introduction and aims of the project**

2. **Publications**

   2.1. Motor Simulation and Ostensive Inferential Communication

   2.2. Motor Simulation and Ostensive-inferential communication: insights and clarifications

   2.3. Experimental Semiotics: A Systematic Categorization of Experimental Studies on the Bootstrapping of Communication Systems

3. **Appendix**

   3.1. Which mindreading for ostensive communication? An event-related potentials study of how the brain processes communicative and informative intentions

4. **Summary and conclusion**

5. **Declarations by all co-authors**

1. Introduction and aims of the project

Language, communication and cognition

The central topics of this thesis are pragmatic competence and, as a specific pragmatic ability, ostensive communication. The central question considered is: does the basic pragmatic competence for ostensive communication rely on advanced, higher-order cognitive processes (thus, being necessarily uniquely human)? Or conversely, can simpler, faster and lower-order cognitive processes suffice to enable basic ostensive communication (thus, making it not necessarily uniquely human)?

To answer this question, I examine the role of pragmatic competence and ostensive communication in human communication and natural language; I propose and defend a model of ostensive communication as anchored in low-level processes such as motor simulation, and demonstrate their importance in the development and acquisition of language; subsequently, through a comprehensive review of experimental semiotics studies I identify the foundations for the emergence of basic communication systems in humans (I also explain the role of experimental semiotics in the investigation of pragmatic competence from an evolutionary point of view); I determine the basic conditions for ostensive communication, from which I conduct an argument that ostensive communication, at least in its basic form, is not an exclusively human capacity but instead is shared with other non-human animals, especially primates. Finally, I flesh out my theoretical position – that ostensive communication works on low-level rather than high-level cognitive processes – with direct empirical evidence acquired via a novel experimental EEG paradigm. In more detail, the papers that are part of the dissertation will examine specific aspects of these topics (Tab. 1). In sum, **the findings of my research project, laid out in this thesis, result in significant implications for research in both pragmatics and language evolution**, which I develop in more detail in Conclusion.

Table 1: Papers that are part of the thesis, with summary and conclusion

Paper	Summary	Conclusion
• Delliponti, A. D. (2022). Motor Simulation and Ostensive-Inferential	The theoretical paper discusses the ostensive model of communication, which emphasizes the distinction	Motor simulation plays a significant role in recognizing both communicative and

Communication. *AVANT. Pismo Awangardy Filozoficzno-Naukowej*, (1), 1-20.	between a speaker's literal meaning and their intended meaning. It explores the role of motor simulation - particularly through mirror neurons - in recognizing communicative and informative intentions in language. It is argued that these processes contribute to language acquisition and may have had implications for the evolution of language.	informative intentions during language acquisition. Specifically, phono-articulatory simulation helps in understanding communicative intentions, while semantic simulation aids in grasping informative intentions. These processes play a role in language development in children and could have influenced the evolutionary transition from gesture to speech in communication.
● Delliponti, A. (2022). Motor Simulation and Ostensive-inferential communication: insights and clarifications. *Theoria et Historia Scientiarum*, *19*, 35-54.	This theoretical paper extends and clarifies the topics of the previous one, discussing the role of motor simulation in language acquisition, specifically focusing on how brain motor areas are activated during the processing of action-related words by infants. It presents a model of ostensive communication, examines mindreading mechanisms in early childhood, and highlights the importance of phono-articulatory and semantic simulations in recognizing communicative intentions.	Motor simulation significantly aids infants in recognizing ostensive cues, which are crucial for language learning. It suggests that low-level processes contribute to understanding communicative and informative intentions, particularly through associative learning during interactions with caregivers. The activation of motor areas in response to verbal stimuli could play a pivotal role in the acquisition of action words and communication.
● Delliponti, A., Raia, R., Sanguedolce, G., Gutowski, A., Pleyer, M., Sibierska, M., ... & Wacewicz, S. (2023). *Experimental semiotics: A systematic categorization of experimental studies on the bootstrapping of communication systems. Biosemiotics, 16 (2), 291–310*.	The paper delivers a structured review of Experimental Semiotics (ES) studies. In ES, human participants communicate without using language; thus, ES can be thought of as 'pure pragmatics' investigation, examining the mechanisms that allow semantics and syntax to emerge. Through categorising available ES studies by their design parameters and results, this article addresses the most	The review identifies two main paradigms in ES: referential games and coordination games. Some of the crucial conclusions - as patterns observed in the studies - are, for example, that cooperation is a fundamental factor for the emergence and establishment of communication systems; or that over time the "signal space" dimension becomes more and more open and continuous.

	basic pragmatic foundations of the emergence of novel communication systems.	
● APPENDIX: Ferretti, F., Angelo D. Delliponti, A., Deriu, V., Chiera, A., Altavilla, D., Nicchiarelli, S., Wacewicz, S., and Adornetti, I. Which mindreading for ostensive communication? An event-related potentials study of how the brain processes communicative and informative intentions (*Under review in Cognitive Science*)	The paper investigates the cognitive processes underlying ostensive communication (OC), specifically the processing of communicative intention (CI) and informative intention (II) as defined by the ostensive model. It addresses the debate between classical and deflationary perspectives on OC, with the classical view considering it uniquely human and requiring high-level mindreading, while the deflationary view suggests basic forms of OC existing in infants and non-human primates, supported by simpler mindreading. The study uses ERPs to examine the time course of processing CI and II expressed through eye contact and gestures.	The ERP analysis revealed findings related to the amplitude of P100, N170, and LC1 (600-800 ms) components. The detection of both communicative and informative intentions appears to occur within a 200-millisecond window. These findings support a deflationary perspective on OC, suggesting that low-level cognitive processes are sufficient for the initial detection of communicative and informative intentions. To my knowledge, the study provides first empirical evidence to inform theoretical debates about the nature of mindreading in OC and its implications for language evolution.

One of the founding aspects of research on the origin and evolution of language, in linguistics and language sciences, is the theoretical debate relating to the cognitive underpinnings of human communication [1, 2, 3]. A crucial aspect of this research path is to identify what are the cognitive requirements that make possible for natural language to have specific features apparently absent in the rest of animal communication. The question is: what is it that led humans to develop a structured, symbolic communication system [4]? One of the most influential positions in recent language evolution research is the *pragmatics-first* approach [1, 3], which argues that a precondition for the emergence of any such system is a certain pragmatic foundation. That is, even before a communication system becomes structured (syntax) and symbolic (semantics), it must already rely on basic pragmatic abilities, such as understanding that the sender intends to "say" something nontrivial, that the receiver will want to "listen" to the sender and try to infer the intended meaning, etc. Even making it clear that my behavior is intended as communicative rather than instrumental (e.g. waving my hand to say goodbye rather than to repel insects) relies on basic pragmatic competence for ostensive communication.

Pragmatic competence and ostensive communication

Pragmatic competence refers to the ability to grasp the intended meaning conveyed through communication, meaning understood as extending beyond the literal meaning of utterances to include context-dependent interpretations. This involves understanding the rules that govern the optimal adaptation of language to both linguistic and extra-linguistic contexts [10]. Essentially, pragmatic competence enables individuals to integrate linguistic and contextual, extra-linguistic information to infer meaning beyond the explicit content of words. Within a pragmatic framework that distinguishes between literal meaning and the speaker's intended meaning, the first serves as a cue rather than the full intended message. As a result, communication often conveys implicit, indirect, or non-literal meanings that go beyond mere linguistic coding. To grasp a speaker's intended meaning, receivers must consider linguistic cues and integrate them with contextual information. Context encompasses spatial, temporal, cognitive, and socio-cultural factors that shape communicative exchanges, including the same linguistic stimuli.

Pragmatic competence therefore refers to a set of cognitive abilities, that allow both the sender and the recipient to interact successfully in a communicative act, an interaction that includes the sender's ability to adequately transfer the information she intends to communicate, and the recipient's ability to adequately understand the information received, namely, grasping the information that the sender intends to communicate in a specific context. One might therefore expect pragmatic competence to play a relevant role in human communication. At the same time, we refer to pragmatics in a specific sense: unless we embrace a "reductionist" conception of what pragmatic competence is (for example, an exclusively human ability), we should expect some continuum on which to place the differences between humans and other animals, or at least attribute to other animals (mostly, primates) the possession of abilities that could be defined as precursors of pragmatics [11].

In this sense, ***ostensive communication*** **can be seen as a specific type of pragmatic competence, something we may have evolved as an optimization process of specific aspects of communication**. The ostensive-inferential model (also: ostensive model, or ostensive communication) explains communication as the process of expressing and recognizing informative and communicative intentions [12]. Informative intention refers to the intention to make the content of a message clear to a recipient; essentially, the information the sender aims to convey, which, in turn, influences the receiver's mental representations. Communicative intention, on the other hand, highlights the act of communication itself. In the classical theorization of the ostensive model, without explicitly signaling the intent to communicate, an informative intention alone is not enough for successful interaction: the production of meaning requires the involvement of mentalistic processes [13].

To better understand, let's imagine this situation: I'm sitting at the counter of a bar and I show the bartender my empty beer glass, after having established eye contact. In this case, through eye contact I express my communicative intention, that is, the very fact that I intend to communicate; instead, by showing the empty glass (e.g. by grabbing the glass and showing it to the bartender) I communicate my informative intention, that is, the fact that I want him to refill my glass. In order to understand this communicative act it is necessary to grasp the communicator's intentions.

Mindreading in ostensive communication

Within the ostensive framework, the sender offers clues to her intentions, while the receiver interprets them. The term "ostensive-inferential" reflects this process: *ostension* refers to presenting communicative cues, while *inference* involves interpreting them. What enables this process is the communicators' ability to reason about their interlocutors' mental states and intentions, a skill part of social cognition. As mentioned, the model relies on contextual factors: in its classical view, the shared beliefs and knowledge between the speaker and listener, as well as their assumptions about each other's understanding. For this reason, another fundamental piece of the ostensive model is the *theory of mind* (ToM), or *mindreading*. Attributing or reasoning about others' mental states necessarily involves some kind of mind-reading activity, which is precisely the ability to attribute desires, beliefs, and intentions to others [14, 15]. Since what characterizes ostensive communication is the expression and recognition of communicative and informative intentions, it seems equally sound to conclude that ToM is a cognitive ability crucial for this specific aspect of pragmatic competence, and more generally, for human pragmatic abilities. This has been highlighted in some parts of the present work [16], and in a way, it is likely that specifically human pragmatic abilities require high-level mindreading, at least in circumscribed conditions. **But what will be questioned in the course of this work, as already pointed out, is that ostensive communication, in its basic form, is an exclusively human capacity** [17, 18].

The problem of ostensive communication arises not only in relation to non-human animals, but also when we try to attribute high-level mindreading capacities to human infants, who are thought to have not yet developed such metarepresentational abilities [19]: a metarepresentation is a higher-order representation that contains or refers to another, lower-order representation. In the previous example of the customer and the bartender, the communicative intention (the intention to make the communicative message explicit) is the higher-order representation that relates to the lower-order, informative intention (the intention to inform the bartender that I want another beer). To address this problem within the ostensive model, there are in principle two possibilities: (i) to assume non-mentalistic stances in ostensive communication, or (ii) to assume the involvement of forms of minimal mindreading: this would mean that **the recognition of communicative and informative intentions does not require embedded representations, but could be the result of low-order representations**. In literature an example of the first (i) case is offered by the model of direct perception [20], while an example of the second (ii) is that of simulative mindreading [21]. The crux of the matter lies in the question: how can infants attribute intentions and mental states to others, and thus communicate ostensively, if they have not yet developed a full-blown ToM? If we assume that ostensive communication is a crucial ability for natural language and human communication, we should also ask how it emerges at different stages of development, unless we consider some developmental leap, in infants, allowing them to transition from non-ostensive to ostensive communicators. Evidence suggests that a gradualist approach is more credible [22]. But then it is also legitimate to ask whether some species of non-human animals, especially non-human primates, may also possess basic capacities for ostensive communication. The question thus shifts from whether a gradualist approach is plausible in analyzing the development of ostensive communication in children, to whether it is legitimate to assume a gradualist approach to its evolution.

Ostensive communication, relevance and language

Before deepening what the role of ostensive communication is in language, we'll introduce the basic notion of the higher-order theoretical approach, that is *relevance theory* [23]. The communicative principle of relevance can be described with the statement that every ostensive stimulus conveys a presumption of its own optimal relevance. With 'presumption of optimal relevance' we mean that:

- The ostensive stimulus is relevant enough to be worth the audience's processing effort.

- It is the most relevant one compatible with communicator’s abilities and preferences.

This principle can be seen as a lawlike generalization about ostensive stimuli and about human communication [24]. In fact, as explained by Wilson & Sperber themselves:

> Knowing your tendency to pick out the most relevant inputs and process them so as to maximize their relevance, I may be able to produce a stimulus which is likely to attract your attention, activate an appropriate set of contextual assumptions and point you toward an intended conclusion [23, pp. 610-611].

It is therefore a cognitive principle, initially formulated to explain a tendency of cognition in general and then adapted to human communication, which is also suitable for describing some of the differences previously observed between animal communication and natural language. But the tendency to optimize the relevance of ostensive stimuli applies to both verbal and nonverbal communication. As in the previous example, if I overtly show the bartender my empty glass in a bar, the relevance of the gesture will almost certainly be interpreted as a request like "Could you refill the glass, please?". Here we have an ostensive stimulus, a context, a communicative intention, and an informative intention. To infer the meaning of the communicative act, the waiter must recognize both the communicative intention, that is, the fact that the customer is overtly showing that she wants to communicate something, and the informative intention, that is, the meaning intended to be conveyed: specifically, the desire for the glass to be full again. Without human cognition's tendency to maximize the relevance of the communicative acts, understanding the meaning that the customer wants to express would be much more complicated, if not impossible. But how do ostensive communication and relevance theory explain the process of production and interpretation at work, not only in nonverbal communication, but also in natural language? We know from previous discussion that pragmatic competence has to do more with understanding the extra-linguistic meaning of a text than with its literal meaning. By ‘extra-linguistic’, we mean the context in which we position the communicative act. It is easy to show that natural language is made up of these two dimensions, namely the literal one relating to the meaning encoded by the language, and the non-literal one referring to the speaker’s meaning, namely her informative intention. If I enter a room where there is an open window saying "it's cold in here", the sentence can be interpreted both as a statement of fact - tthat it is indeed cold in the room - but also as a request to close the window. It can also be shown that here, too, we have an ostensive stimulus (the sentence), a context, a communicative intention, and an informative intention. Decoding the literal meaning of the sentence is therefore not enough to understand the speaker's informative intention: we must infer the intention itself.

Traditional and pragmatics-first approaches to language emergence

Given what has been said so far, a possible conclusion is that relevance theory and ostensive communication play a fundamental role in human communication and language. In taking this position, we would be supporting a theoretical model asserting a pragmatics-first perspective, according to which the pragmatics of communication, including nonverbal communication, is the foundation of natural language, explaining the discontinuities found with animal communication [3, 11]. This approach also differs from the Chomsky's classical universal grammar framework, which focuses on syntax [2], according to which the main property of language is that of generating an infinite number of sentences starting from a finite set of items and rules, and more generally, from approaches that focus on the semantic dimension of language. This means that syntax and semantics alone would not be sufficient to explain the differences between natural language and animal communication, or maybe would not explain those differences at all. Instead these differences are explainable by focusing on pragmatics, and specifically, on the difference between humans' and non-human animals' pragmatic abilities. **Relevance theory and ostensive communication would then account for much of the observed differences between animal communication and natural language** [13].

Some differences between animal communication and natural language

An influential attempt to delineate the differences between animal communication and human language were Hockett's *design features* [5, 6, 7]. However, the design features have been criticized by several scholars - from various perspectives - and there have been attempts to update them [8]. Furthermore, these features have a very broad scope, while what we want to highlight here are specific properties more directly connected to a core of cognitive abilities related to pragmatics.

Scott-Phillips [9] proposed some differences between human language and the communication systems identified and documented in various nonhuman animals. These are:

- *Spontaneous versus volitional*. Nonhuman vocal communication resembles sighs, grunts, laughter, and other similar spontaneous activity, but it's different from deliberate forms of expression like those we use in natural language.

- *Finitude versus infinitude*. Natural language has the potential to generate a countably infinite number of sentences, but there is currently no evidence of such limitless complexity in nonhuman communication.

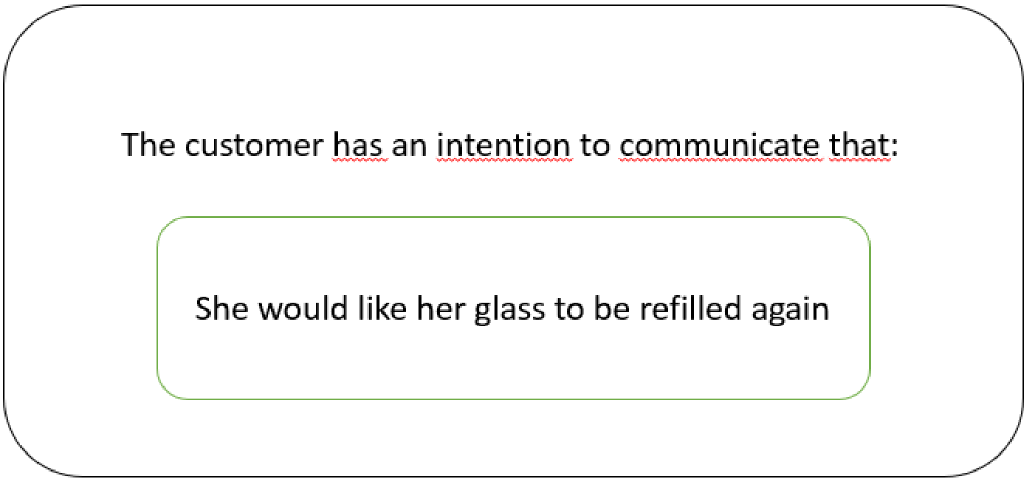
- *Literalism versus contextualism*. Words and other linguistic elements are intertwined with human cognitive abilities for communication in ways that have no clear parallels in other species.

- *Narrow domains versus virtual open-endedness*. Humans seem to possess an unlimited range of communicative goals, unlike other species. These goals appear to contradict some fundamental principles of evolved communication systems.

The role of relevance and ostensive communication in natural language

But how would relevance theory and ostensive communication account for the differences highlighted by Scott-Phillips? Let's focus on the first dissimilarity: *spontaneous vs. volitional* vocalizations. Humans use their voices in two main ways: spontaneous, like laughter or screams, which arise automatically in response to emotions, and volitional, which are intentional. While laughter can be faked, genuine spontaneous vocalizations resemble nonhuman primate calls more than language. Evidence includes acoustic properties and brain structures, suggesting an evolutionary link between human spontaneous vocalizations and primate calls, distinct from language. However, the presence of intentional communication in nonhuman primates is debated, as some apes seem capable of voluntary communication. What sets human language apart is ostensive communication, the overt expression and recognition of communicative intent. This voluntary aspect is largely absent in animal communication due to a lack of metarepresentation (e.g., understanding "I know you intend to communicate X") (Fig. 1), which is tied to Theory of Mind. Yet, we will question whether ostensive communication truly requires such advanced mindreading, as suggested by Moore [28].

Figure 1: A case of metarepresentation, with a representation embedded in another one



Another key difference between animal communication and natural language is *literalism vs. contextualism*. Many studies assume words have fixed meanings (“presumption of literalness”), but research shows linguistic comprehension is highly context-dependent, adjusting dynamically [29, 30]. For example, “It's cold in here” might imply a request to close a window. Studies reveal that nonliteral meanings (e.g., metaphor, irony) are processed as fast as literal ones, contradicting the idea of a separate pragmatic step. Instead, comprehension involves real-time interaction between linguistic decoding and context [31]. This parallel processing is unique to humans; no evidence suggests nonhuman animals interpret communication this way. While animals, especially apes, can adapt gestures to context, the depth of contextuality in human language is unparalleled. This ability seems tied to the principle of relevance: ostensive stimuli imply optimal relevance. However, cognitive differences between humans and animals likely extend beyond communication, requiring exploration of both the degree and type of cognitive abilities involved [4].

As for the *finitude vs. infinitude* dissimilarity, human language uniquely generates infinite sentences through recursion, or Merge [2], which hierarchically combines syntactic elements. Animal communication, by contrast, is limited to zero-merge or one-merge systems, meaning signals are either separate or simply combined. No species uses two-merge or higher, essential for hierarchical structures. Even Kanzi, an ape with linguistic skills, struggles with hierarchical grammar [33]. Some suggest animals may have complex syntax that we haven't observed or lack cognitive capacity for higher-order merging. However, no evidence shows nonhuman communication has the open-endedness of human language [34]. This fundamental cognitive difference may stem from working memory limits or, as Scott-Phillips suggests, the smaller communicative range in animals makes complex merging unnecessary [9]. Evolutionary pressures likely expanded human communicative domains, with ostensive communication potentially playing a role. This connects to the next dissimilarity: *narrow domains vs. virtual open-endedness*. As Scott-Phillips argues, infinite merging isn't needed if the communicative range is small.

This fourth point analyzed by Scott-Phillips refers to the fact that human communication is uniquely open-ended, allowing for infinite topics and methods, including language, gestures, and improvised signals. In contrast, non-human communication is domain-specific, tied to survival needs (like bees signaling flower locations). Evolution favors stable, mutually beneficial communication, which typically limits its scope, yet humans defy this constraint by discussing abstract ideas and future events. Some theories attribute this to our metapsychological abilities [35], while others emphasize language's combinatorial nature. Regardless of the explanation, no known animal system matches the full flexibility of human communication. This open-endedness, an anomaly from an evolutionary perspective, could allow for infinite possibilities of deception, while animal communication remains narrow. Various explanations have been given for this peculiarity [36], including evidence that reputation plays a fundamental role in human communication [42]. Ostensive communication - making one's communicative intention explicit - activates higher-order mentalization in the receiver, leading to inferences about the sender's intentionality and commitment, a precursor to a sense of personal responsibility. These mechanisms, linked to high-level mindreading, enable a vigilance system that reinforces the speaker's reputation, thus stabilizing the evolutionary strategy by limiting deception through social exclusion or punishment.

Description of the content

In the course of this work, various aspects concerning the foundations of pragmatics and ostensive communication will be addressed. We will first present some proposals on low-level mechanisms involved in ostensive communication, specifically their role during human development. These theoretical models are interconnected with embodied approaches: we will analyze how certain aspects of ostensive communication may be framed as originating in motor simulation, and examine the role of perception and motor resonance in the development of ostensive communication. **These papers, which focus on low-level cognitive processes, are consistent with the thesis's broader attempt to reassess ostensive communication as a pragmatic competence that, rather than being unique in the natural world** (i.e., exclusive to humans), **could be considered as a set of cognitive properties partially present in other animals**, such as non-human primates [37]. This is the hypothesis of evolutionary continuity, examined here through the study of embodied models such as motor simulation and how these capacities manifest in other species. The aim is not to claim that there are no differences between ostensive communication in humans and other animals, but that **the foundations of ostensive communication, or basic forms of it, could hypothetically be identified in non-human animals, especially apes**.

In this section, the papers that make up the thesis are briefly described, along with their role in explaining the thesis' central arguments. In the first paper, "Motor Simulation and Ostensive-Inferential Communication" [16], it is argued that the ostensive model has an embodied basis, particularly emphasizing the role of motor simulation in recognizing informative and communicative intentions. It explores a possible link to the mirror system hypothesis, suggesting that language development might be intertwined with gestural communication and with the mechanisms related to the mirror neuron system. This system is crucial, as it enables individuals to grasp others' actions and intentions through a process of internal simulation. The theoretical proposal is that motor simulation - specifically phono-articulatory (speech-related) and semantic simulation - plays a crucial role in recognizing intentions during language acquisition. Within this framework, an embodied approach to understanding the origins of human communication is suggested, arguing that motor simulation provides a foundation for recognizing communicative and informative intentions. Motor simulation may have played a significant role in the biological evolution of language, particularly in the context of ostensive-inferential communication. This view is contrasted with the Universal Grammar approach, which emphasizes innate structural rules independent of pragmatic competence.

The second paper, "Motor Simulation and Ostensive-inferential communication: insights and clarifications" [43], extends the earlier exploration of embodied mechanisms in language acquisition, specifically focusing on how motor simulation facilitates the recognition of communicative intentions. This paper builds upon the previous argument that motor areas play a role in infants' recognition of communicative and informative intentions. Three key expansions of the initial model are proposed: firstly, a connection between the motor simulation model to the specifics of infant-caregiver interactions in speech perception, particularly emphasizing the role of "baby talk." It posits that the unique prosodic features of infant-directed speech - pitch, amplitude fluctuation, and speed - serve as ostensive cues. This "phono-articulatory simulation" helps infants to recognize that communication is being directed at them. It leverages infants' innate sensitivity to speech to facilitate the recognition of (linguistic) communicative intentions. Secondly, an explanation of the development of networks between speech areas and the motor cortex. This is outlined by explaining how semantic simulation occurs, and showing evidence regarding how the motor cortex is elicited by the content of action words. This connection arises through associative learning, playing a crucial role in acquiring action-related words (nouns, verbs, adjectives). Thirdly, an examination of how different **mindreading models** are consistent with the embodied simulation theory and with childrens' cognitive abilities: the aim is to show the compatibility of motor simulation with existing frameworks of intention recognition. In short, the paper deepens and clarifies the topics of the previous one by grounding them in the literature on infant-caregiver interaction, and outlining the neural mechanisms underlying motor simulation within different mindreading models. It provides a more detailed and nuanced explanation of how motor simulation contributes to language acquisition and intention recognition in early development.

The fundamentals of ostensive communication and pragmatics

As mentioned above, ostensive communication can be seen as a specific skill that is part of the broader pragmatic competence. At the same time, another aim of the thesis is to investigate what are the foundations of pragmatics, in the most primary sense of understanding meaning in relation to context. The third paper, "Experimental Semiotics: A Systematic Categorization" [44], offers a complementary perspective to the previous discussions on embodied communication. Embodied here in a specific, slightly different sense: investigating what are **the foundations of pragmatic competence**, from the perspective of experimental studies concerning **the process of creating the relation between signs and their interpreters as biological, psychological, and social agents**. While previous works explored language acquisition through motor simulation and ostensive communication, this paper examines the experimental literature on how novel communication systems emerge without the possibility to rely on pre-existing communication systems, such as natural language. While the motor simulation framework is about the embodied mechanisms that enable infants to recognize communicative and informative intentions, *experimental semiotics* investigates **the bootstrapping of communication from scratch, examining how individuals create meaning and establish conventions**. An important contribution of this paper is its systematic review and categorization of experimental semiotics studies, also identifying types and properties of experiments about emergent communication systems, such as:

- **Type of Game:** Referential and coordination games, each requiring different communicative strategies.
- **Modality of Communication:** Gestures, drawings, vocalizations, and so on, the diverse channels humans can use to convey meaning.
- **Presence of Turn-Taking:** Investigating how interactional dynamics influence the development of communication.
- **Transmission Methods:** Vertical vs. horizontal, showing how communication systems are transmitted and how they evolve.

Experimental semiotics is a distinctly notable area of language evolution research because it focuses on the experimental making of meaning from scratch, a way of investigating how meaning can emerge when agents need to coordinate to establish communication strategies, with the means they have, in order to understand each other. In this way it is also possible to investigate the bases of pragmatics from an evolutionary perspective, since one of its goals is to examine which are the modalities and constraints that make certain features of human communication and language emerge. Actually, an interesting aspect is linked to the fact that different means of communication (e.g., gestures or symbols) show properties similar to natural language, for example as in the case of arbitrariness or systematicity [38, 39]. Furthermore, pragmatics here is also understood in an ecological sense, that is, another recurring element of this field of study is that it is not only cognitive constraints that determine or favor the emergence of the properties of human communication systems, but also environmental constraints. This has also got a pragmatics orientation at an evolutionary level, as it refers to the context in which a communication system can emerge. Therefore, the evolution of human pragmatics skills must always be traced back to the adaptation demands of our species, hence to the environment in which we evolved and to the features that have brought out the properties - including those we've spoken about previously - that are typical of natural language. Another interesting aspect of experimental semiotics is the exploration of multimodal communication, multimodality that is in fact one of the main themes related to the discussion on the origin of language, whether it is postulated that language originates in gestures, in pantomime, or in multimodal communication [40]. As already mentioned, different means show similar properties: **experimental semiotics is then a suitable research field to investigate these means and their properties, in line with the hypothesis that pragmatics is the foundation of communicative modalities so different and yet quite similar in their deep structure**. Finally, with experimental semiotics it is possible to investigate the basis of how ostensive communication emerges [41]. Once again, even if we assume that ostensive communication is composed of a set of skills whose core is shared between humans and other species, it is anyway to pragmatic skills - that go beyond speech - that we must look to understand the origins of language.

This work also suggests possible research paths in experimental semiotics starting from ignored aspects or from some of the most interesting perspectives that emerge from a careful analysis of the literature.

Which mindreading for ostensive communication?

In this section I summarize in more depth the content of the last paper (under review), present in the Appendix, entitled "Which mindreading for ostensive communication? An event-related potential study of how the brain processes communicative and informative intentions" [17]. This research delves into the question of how our brains process the intentions behind communication. Specifically, it investigates the cognitive mechanisms involved in ostensive communication". As partially anticipated previously, the traditional view suggests that ostensive communication, with its two layers of intention (communicative and informative), is uniquely human and relies on sophisticated, *recursive*, mindreading abilities. However, there is a more recent perspective according to which even infants and primates might rely on simpler versions of ostensive communication. The question is, what kind of mindreading is actually in basic ostensive communication. Are we talking about recursive, inferential reasoning? Or is it something way more basic, some low-level and automatic process?

In order to investigate this, we resorted to electroencephalography (EEG), a method used to record electrical activity in the brain: metal electrodes are placed on the scalp, detecting electrical activity that results from neurons. In addition to this, signals are then amplified and recorded. EEG is useful for studying brain activity through a high temporal resolution, so it’s a very good method for examining the timing of cognitive processes. Starting from this, EEG data was analyzed to identify event-related potentials (ERPs), that are specific components in the EEG recordings, time-locked to events or stimuli. When your brain processes something specific (like seeing someone make eye contact), a particular pattern of electrical activity occurs, and ERPs are suitable for isolating and studying them, providing insights into the timing and nature of the underlying cognitive processes. Different ERP components are related to different types of information processing.

The experiment looked at how people process communicative and informative intentions, using stimuli like eye contact and gestures. What showed up in the data were these specific ERPs:

- P100 and N170: These are early components, ranging in the first 200 milliseconds. The study found that both communicative and informative intentions elicited activity related to these early components.
- LC1 (600-800 ms): This component is detected later, reflecting processes occurring several hundred milliseconds after the initial stimulus.

Finding such early components suggests that the brain processes both communicative and informative intentions quite early, i.e. **recognizing communicative and informative intentions in the framework of basic ostensive communication is related to fast, automatic processes that rely on low-level mechanisms rather than high-level, inferential reasoning**. This lends support to the deflationary perspective on ostensive communication, suggesting that basic forms of mindreading may be sufficient for these initial stages of communication.

These findings have several implications:

- They are important for the debate around relevance theory.
- They sheds light on the neural timeline of basic mindreading (or direct perception), pinpointing when different aspects of intention are processed.
- They help us understand the kind of mindreading involved when we're dealing with basic ostensive communication.
- The results align with research showing that basic forms of ostensive communication emerge early in development.
- The data are consistent with the idea that basic forms of ostensive communication might also be present in non-human primates.
- The findings support a pantomimic scenario for the origin of language, suggesting that early communication may have relied on gestures and shared intentions.

My contribution: as acknowledged by my shared first-authorship of this paper, I contributed to all the stages of this project: from the design of the project to data collection, data analysis, and writing up the paper. My involvement spanned from the initial conceptualization of the experiment to the interpretation of the results and the crafting of the final manuscript.

References

1. Bar-On, D. (2021). How to do things with nonwords: pragmatics, biosemantics, and origins of language in animal communication. *Biology & Philosophy, 36*(6), 50.

2. Berwick, R. C. (2016). Why only us: Language and evolution.

3. Moore, R. (2017). Pragmatics-first approaches to the evolution of language. *Psychological Inquiry, 28*(2-3), 206-210.

4. Delliponti, A. D., Zhang, E. Q., Ng, Y. Y., & Pleyer, M. (2024). The cognitive requirements for developing a multimodal communication system: Evidence from experimental semiotics and comparative cognition. The Evolution of Language Conferences.

5. Hockett, C. F. (1959). Animal" languages" and human language. Human Biology, 31(1), 32-39.

6. Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. Scientific American, 203(3), 88-97.

7. Hockett, C. F. (1963). The problem of universals in language. Universals of language, 2, 1-29.

8. Pleyer, M., & Hartmann, S. (2024). Cognitive Linguistics and Language Evolution. Elements in Cognitive Linguistics.

9. Scott-Phillips, T., & Heintz, C. (2023). Animal communication in linguistic and cognitive perspective. Annual Review of Linguistics, 9(1), 93-111.

10. Bambini, V. (2017). Il cervello pragmatico. Rome: Carocci.

11. Bar-On, D. (2024). ‘Pragmatics First’: Animal Communication and the Evolution of Language. Review of Philosophy and Psychology, 1-28.

12. Sperber, D. (1986). Relevance: Communication and cognition.

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

14. Apperly, I. (2010). Mindreaders: the cognitive basis of" theory of mind". Psychology Press.

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

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

17. Ferretti, F., Delliponti. A. D., Deriu, V., Chiera, A., Altavilla, D., Nicchiarelli, S., Wacewicz, S., and Adornetti, I. Which mindreading for ostensive communication? An Event-Related Potentials (ERPs) study of how the brain processes communicative and informative intentions. (Under review).

18. Sperber, D., & Wilson, D. (2024). Rethinking ostensive communication in an evolutionary, comparative, and developmental perspective. PsyArXiv: https://doi. org/10.31234/osf. io/zp3fx.

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

20. Gallagher, S. (2008). Direct perception in the intersubjective context. Consciousness and cognition, 17(2), 535-543.

21. Shanton, K., & Goldman, A. (2010). Simulation theory. Wiley Interdisciplinary Reviews: Cognitive Science, 1(4), 527-538.

22. Csibra, G. (2010). Recognizing communicative intentions in infancy. Mind & language, 25(2), 141-168.

23. Wilson, D., & Sperber, D. (2006). Relevance theory. The handbook of pragmatics, 606-632.

24. Scott-Phillips, T. (2024). The Communicative Principle of Relevance. Current Directions in Psychological Science, 33(6), 371-377.

25. Bryant, G. A., & Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. Evolution and Human Behavior, 35(4), 327-335.

26. Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. Behavioral and Brain Sciences, 37(6), 529-546.

27. De Waal, F. (2016). Are we smart enough to know how smart animals are?. WW Norton & Company.

28. Moore, R. (2016). Meaning and ostension in great ape gestural communication. Animal Cognition, 19(1), 223-231.

29. Wilson, D. (2003). Relevance and lexical pragmatics. Italian Journal of Linguistics, 15, 273-292.

30. Gibbs Jr, R. W., & Colston, H. L. (2012). Interpreting figurative meaning. Cambridge University Press.

31. Wilson, D., & Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts.

32. Bohn, M., Liebal, K., Oña, L., & Tessler, M. H. (2022). Great ape communication as contextual social inference: a computational modelling perspective. Philosophical Transactions of the Royal Society B, 377(1859), 20210096.

33. Truswell, R. (2017). Dendrophobia in bonobo comprehension of spoken English. Mind & Language, 32(4), 395-415.

34. Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How could language have evolved?. PLoS biology, 12(8), e1001934.

35. Heintz, C., & Scott-Phillips, T. (2023). Expression unleashed: The evolutionary and cognitive foundations of human communication. Behavioral and brain sciences, 46, e1.

36. Scott-Phillips, T. C. (2008). On the correct application of animal signalling theory to human communication. In The evolution of language (pp. 275-282).

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

38. Caldwell, C. A., & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties.

39. Nölle, J., Staib, M., Fusaroli, R., & Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. Cognition, 181, 93-104.

40. Zlatev, J., Wacewicz, S., Zywiczynski, P., & van de Weijer, J. (2017). Multimodal-first or pantomime-first? Communicating events through pantomime with and without vocalization. Interaction Studies, 18(3), 465-488.

41. Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. (2009). Signalling signalhood and the emergence of communication. Cognition, 113(2), 226-233.

42. Giardini, F., Fitneva, S. A., & Tamm, A. (2019). " Someone told me": Preemptive reputation protection in communication. PloS one, 14(4), e0200883.

43. Delliponti, A. (2022). Motor Simulation and Ostensive-inferential communication: insights and clarifications. *Theoria et Historia Scientiarum*, *19*, 35-54.

44. Delliponti, A., Raia, R., Sanguedolce, G., Gutowski, A., Pleyer, M., Sibierska, M., ... & Wacewicz, S. (2023). Experimental semiotics: A systematic categorization of experimental studies on the bootstrapping of communication systems. *Biosemiotics*, *16*(2), 291-310.

2. Publications

2.1. Motor Simulation and Ostensive-Inferential Communication

AVANT, Vol. XIII, No. 1
ISSN: 2082-6710 avant.edu.pl/en
DOI: 10.26913/avant.202204

Motor Simulation and Ostensive-Inferential Communication

Angelo D. Delliponti
Academia Copernicana,
Nicolaus Copernicus University
angelo.d.delliponti@gmail.com

Received 9 September 2021; accepted 19 April 2022; published 17 June 2022.

Abstract

The ostensive-inferential model is a model of communication, an alternative to the code model of communication, based on pragmatic competence: it explains human communication in terms of expression and recognition of informative and communicative intentions, founding comprehension on the distinction between literal meaning and the speaker's meaning. Through informative intentions we try to make evident the content of a message to a receiver, or to make evident what we want to communicate to him/her: communicative intentions are used to make evident the very fact that we intend to communicate. One hypothesis is that ostensive-inferential communication is what makes human language possible. Since an extensive literature has highlighted the role of the Theory of Mind in ostensive-inferential communication, this hypothesis fits with the idea that a mechanism for mentalizing underlies human communication. The aim of the present paper is to stress the role of lower-level mechanisms, specifically of motor simulation, in the recognition of informative and communicative intentions, in order to outline an embodied account of ostensive communication. Specifically, the hypothesis is that this process is involved in language acquisition during development, and that it plays a role in the associative learning process involved in language acquisition during childhood. To this aim, in future research it may be useful to test the involvement of motor simulation (specifically, phono-articulatory and semantic) in the recognition of informative and communicative intentions in toddlers. Since some models of language evolution focus on the role of motor simulation, a supplementary goal is to deepen its role in the biological evolution of language, focusing on the specific link between motor simulation and intentions in the framework of ostensive-inferential model.

Keywords: ostensive communication; motor simulation; communicative intention; informative intention; embodiment; language acquisition

1. Introduction

In this article I argue in favour of the hypothesis that ostensive-inferential communication has an embodied basis, stressing in particular the importance of lower-level mechanisms such as motor simulation in the recognition of informative and communicative intentions. I begin by explaining what ostensive-inferential communication is and what are the theoretical principles and the experimental evidence that would make it possible to claim that it has an embodied basis. In doing this, I present a theory according to which the origin of human communication is anchored to gestural communication, which may have guided vocal communication throughout the evolution of language: it is the Mirror System Hypothesis (MSH) (Arbib, 2012). This allows me to show the existing link between ostensive-inferential communication and language: in fact, mirror neurons are involved in the recognition of intentions (Gallese, 2007) and in the processing of words with action content (Pulvermuller, 2005). I then review studies that show how communication and language, in production and understanding, involve areas of the brain dedicated to motor processing (Buccino et al., 2001; Fadiga et al., 2002; Hauk et al., 2004; Martin et al., 1996).

Building on those foundations, I hypothesize that there are two subsets of motor simulation—i.e., the reactivation of sensorimotor patterns, extrapolated from their motor functions and exploited in cognitive processes different from those for which they evolved or during which they formed (Borghi & Caruana, 2016)—involved in recognizing informative and communicative intentions during language acquisition: they are phono-articulatory simulation and semantic simulation. Phono-articulatory simulation, which occurs with activation of motor cortex areas involved in speech production, is involved in the recognition of communicative intentions, while semantic simulation, which occurs with activation of motor cortex areas involved in processing action content words, has a role in the recognition of informative intentions. What I want to emphasize here is that these mechanisms play an important role in the acquisition of language during development. Consequently, it is also possible to hypothesize (in a completely speculative way) that motor simulation has a more general role in the recognition of communicative and informative intentions in linguistic communication.

2. Pragmatic competence and ostensive-inferential communication

With pragmatic competence we mean the ability to understand the message conveyed by the utterances in the course of communication; it refers not only to the literal meaning, but mainly to the meaning linked to the context, i.e. the knowledge of the rules of optimal adaptation of a language to the linguistic and extra-linguistic context within which communication takes place (Bambini, 2017). Pragmatic competence is, therefore, the ability to integrate linguistic information with contextual information, in order to understand the meaning of communication beyond the strictly literal level. Within a pragmatic approach based on the distinction between literal meaning and speaker's meaning, the former is a hint of the message: it is therefore possible to transmit messages whose meaning is not to be found only in the code, because the content is often implicit, indirect or non-literal.

To understand the meaning of the speaker, it is thus necessary to take into account the linguistic clue and to integrate linguistic material with the context. The notion of context is not uncontroversial. In my perspective, it can be described as the set of space-time, and cognitive and socio-cultural coordinates in which communicative exchanges take place, including the linguistic material of the discourse (Bambini, 2017). The ostensive-inferential model (or ostensive communication), a communication theory alternative to the code model (Shannon & Weaver, 1949), explains human communication in terms of expression and recognition of informative and communicative intentions (Sperber & Wilson, 1986). But what is a communicative intention and what is an informative intention? With informative intentions we try to make evident the content of a message to a receiver, or what we want to communicate to her: the content of an informative intention is the information provided to the interlocutor, and this information corresponds to the changes that the sender intends to produce in the mental representations of the receiver. With the communicative intentions we want to make evident the very fact that we intend to communicate; furthermore, if the expression of an informative intention is not accompanied by the expression of a communicative intention, communication itself fails (Scott-Phillips, 2015). According to the theoretical framework of ostensive-inferential model, the sender provides hints of his intentions and the receiver interprets them: in fact, the meaning of "ostensive-inferential" is precisely this, that is, ostension as an offer of clues and inference as an interpretation of the clues (Scott-Phillips, 2015). What makes the process possible is the fact that whoever communicates can reason about the intentions and mental states of the interlocutor: the intention of the person who produces ostensive stimuli is, in fact, to modify the mental states of the receiver, and not simply, as in the code model, to send a message to be decoded. So this depends on contextual factors, that is, on the beliefs and knowledge that a speaker has of a listener's beliefs and knowledge, and vice versa (Scott-Phillips, 2015).

To better clarify the difference between communicative intention and informative intention, I will take an example directly from Scott-Phillips (2015):

> I am in a coffee shop, I catch the eye of the waiter, and I tilt my coffee cup in a particular, somewhat stylized way. The waiter then comes over and refills my cup. Here, I have an informative intention that the waiter understands that I would like a refill. And so on. The content of an informative intention is, in colloquial terms, the information that it provides. More specifically, it is the changes that the signaler wants to make to the receiver's mental representations. [...] The tilt of my coffee cup expresses my informative intention, but it also expresses something just as important: the very fact that I wish to communicate with the waiter at all. [...] How does the tilt reveal to the waiter that it is a signal? [...] I must also make it clear to the waiter that I am trying to communicate with him at all. My intention to do this—that is, my intention to create in my audience a representation of the fact that I have an informative intention—is called a *communicative intention*. This intention is expressed when I establish eye contact with the waiter and tilt my cup in a particular, *ostensive*, way (pp. 35-36).

3. Pragmatics and the Mirror system hypothesis (MSH)

Scott-Phillips (2015) states that ostensive communication comes before language and that transition to language became possible only after the birth of ostensive communication. „What made ostensive communication possible ultimately made language possible too" (p. 134). Said {thus?} this, the mechanism that makes ostensive communication possible is the cognitive module of Theory of Mind, that is the ability to represent others' mental states and to reason about their thoughts (Premack & Woodruff, 1978; Apperly, 2011; Byom & Mutlu, 2013): the reason why mindreading is considered fundamental to pragmatic competence is that, in the act of communicating, it is important to know certain aspects of the mental dimension of our interlocutor in order to understand his intentions (Scott-Phillips, 2015). The theory of Mind is therefore an essential starting point and, despite the attempts to replace it with various theoretical proposals (see for example Gallese, 2007), it would seem difficult to think of a social cognition or, specifically, a pragmatic competence without a Theory of Mind.

Nevertheless, here—as I will show later—I argue that it is not possible to understand ostensive communication without reference to any embodied foundation, or without integrating it with the mechanism of motor simulation (Borghi, Caruana, 2016; Gallese, 2007), which I assume to be involved in the recognition of communicative and informative intentions. The idea is that motor simulation, through mirror neurons, has made the biological evolution of human language possible, and that at its base there is the expression and recognition of informative and communicative intentions.

Motor simulation has been used in the evolutionary literature for the definition of different models of language, one of which was proposed by Arbib (2012) through the Mirror System Hypothesis (MSH), an approach that attempts to outline the evolution of language by comparing the systems of praxis and communication of human and non-human primates (Arbib & Rizzolatti, 1997; Rizzolatti & Arbib, 1998), later developed into Cognitive Neuroprimatology (CNP)—(Arbib et al., 2018). In summary, the hypothesis draws on the findings by Poizner et al. (1987), according to which deaf people's lesions in the Broca's area induce, with respect to sign language, a form of aphasia similar, in its outcome, to that of spoken language in subjects with intact hearing. Hence, the hypothesis is that mirror neurons could be the basis for language parity, namely the fact that listeners are able to grasp the speaker's meaning thanks to a system that has a mirror mechanism for gestures at its base, with manual gestures that may have guided vocal gestures throughout the evolution of language. MSH postulates recognition and imitation of complex action as a foundation of the emergence of the language-ready brain (Arbib, 2013): this is compatible with the idea that ostensive communication comes before language and that the transition to language became possible only after the appearance of ostensive communication. Language parity and motor simulation are connected with the neural exploitation hypothesis: the main assumption is that the key aspects of human social cognition are supported by neural exploitation, i.e. an adaptation of the brain mechanisms of sensorimotor integration in order to use them for new purposes concerning thinking and language, and at the same time retain their original functions (Gallese, 2003; Gallese & Lakoff, 2005).

MSH is a gestural hypothesis on the origin of language. There is an extensive literature that places gestural communication as a starting point in the evolution that led to vocal language (Hewes, 1973; Arbib, 2012; Armstrong et al., 1995; Corballis, 2002; Stokoe, 2002; Tomasello, 2008), although, as highlighted by Zywiczynski et al. (2017), in recent years the hypothesis of the original multimodality of proto-language has made its way (Kendon, 2011; McNeill, 2012; Sandler, 2013). There is therefore a vast literature in support of MSH, albeit I do not wish to deny here the possibility of a multimodal origin of human communication and language. I only emphasize that, since the two communication systems—vocal and gestural—for some researchers (McNeill, 2012), are integrated to the point of being part of a single cognitive system (Zywiczynski et al., 2017), I here support the possibility of the compatibility of MSH with the multimodal scenario. However, investigating this aspect goes beyond the scope of this article.

By the way, the present proposal differs from previous models of language evolution because it focuses on the link between motor simulation and intentions, in the framework of ostensive-inferential communication. At the same time, through the latter I intend to distance myself from the syntactic approach of the Universal Grammar (UG): I actually believe that UG (Chomsky, 1957), unlike ostensive communication (which rests on pragmatics), is unable to explain the huge creativity and flexibility of human communication. Indeed UG, which is fully compatible with the code model, states that the basis of language is a set of innate structural rules that evolved completely independent of any pragmatic competence.

But what is the other evidence in favour of the neural exploitation hypothesis and of language parity? Some researchers (Masataka, 2001; Gentilucci et al., 2004a; Bernardis & Gentilucci, 2006) showed that there is a close relationship between the development of both oral and manual motor skills. Among the proposals supporting this thesis is the idea that speech production and manual gestures related to speech can be considered as results of the same process (Goldin-Meadow, 1999); the fact that babbling in 6-8 month old babies is accompanied by rhythmic hand movements (Masataka, 2001); or that children born to deaf parents show hand movement with a rhythm similar to that of babbling (Gallese, 2007). There is also a close relationship between linguistic articulation and manual gestures linked to oral language even in adulthood: in a study by Gentilucci et al. (2004a), participants had to either grasp and bring to the mouth fruits of different sizes such as a cherry or an apple, or observe the same actions performed by someone else, while simultaneously pronouncing the syllable /ba/. What was highlighted is that the second formant of the vowel a, linked to the position of the tongue, increased when they performed or observed the act of bringing the apple to the mouth (or its pantomime), which was the largest object, compared to the case in which the same operations were done with the cherry: this means that the execution/observation influenced the speech production, and that the system involved shares the premotor neural circuits involved in the control of arm/hand actions. Furthermore, it is possible that language production comes precisely from those same mechanisms (Gallese, 2007). Another study is the one of Bernardis and Gentilucci (2006), in which participants had to pronounce words (such as "hello" or "stop"), make communicative arm gestures with the same meaning, or emit the two signals at the same time: results showed how the vocal spectrum of the words was reinforced by the simultaneous execution of the gesture with the same meaning (the second formant) compared to when the words were pronounced alone. The same thing did not happen when the words were meaningless. Saying the words tended rather to inhibit a simultaneous execution of the gesture, and even in this case the effect was not visible with pseudo-words. Subsequently it was found that the reinforcement effect was also present when words were pronounced in response to listening to them and to the simultaneous observation of corresponding gesture by a third person. These results show that „spoken words and symbolic communicative gestures are coded as a single signal by a single communication system within the premotor cortex" (Gallese & Glenberg, 2012, p. 36). Other studies confirm the involvement of Broca area (Gentilucci et al., 2006): since the region contains mirror neurons, it is very likely that the communicative meaning of gestures is merged, through motor simulation, with the articulation of the sounds required to express them in words.

4. Understanding intentions

I claimed that ostensive-inferential communication explains human communication in terms of expression and recognition of informative and communicative intentions (Sperber & Wilson, 1986). But what does it mean to understand the intentions behind someone else's actions? According to Gallese (2007), understanding the reason for performing a certain act, such as grabbing a cup, means detecting the goal of the next imminent and not yet completed act, for example, bringing the cup to the mouth. At the basis of this theorization there was an experiment carried out with functional magnetic resonance imaging (fMRI) (Iacoboni et al., 2005). Volunteers observed three types of stimuli: actions such as a grasping hand without context; the context, like a scene with objects; and a grasping hand inserted in some context. The observation of motor acts within a context, compared with the other two experimental conditions, produced a significant increase in the signal in the posterior part of the inferior frontal gyrus and the ventral premotor cortex, correlated with the actions of the hands. Therefore, according to Gallese (2007), the premotor mirror areas, active both during the execution and during the observation of the movements, are not only involved in the recognition of the action, but also in understanding the reason for an action, or rather the intention of its underlying motives. It would thus be the mirror system to make possible this mechanism through the automatic activation of motor simulation.

Another study (Fogassi et al., 2005) found a class of mirror neurons in the parietal area whose activation during the observation of an act, such as grasping an object, is conditioned by the kind of subsequent act not yet detected, for example bringing the object to the mouth, thus specifying the overall intention of the action; these neurons are activated only in reference to the execution/observation of motor acts linked to a specific action, but aimed to a more distal goal: this neuronal activation occurs in a monkey before the execution/observation of the movement linked to the distal goal. According to Gallese (2007), this means that in addition to target recognition, mirror neurons allow the observing monkey to perform a targeted act (for example, bringing an object to the mouth rather than placing it in a container) to predict what the agent is about to do, thus understanding the overall intention of the action. This mechanism found in non-human primates could be the basis of the most sophisticated forms of understanding intentions typical of our species.

Mirror neurons could therefore play an important role in the recognition of intentions. It could be the recognition of the speaker's intentions—through mirror neurons—at the basis of our communicative ability, and specifically, the recognition of communicative and informative intentions as the basis for the evolution of language. In fact, we will see how mirror neurons could have made it possible to move from the recognition of intentions for primordial communicative purposes (at the beginning presumably, as already seen, in the form of manual gestures) to language, and we will see it by showing that mirror neurons are also involved in recognizing intentions in language.

Before continuing, however, it is right to make a clarification on the role that mirror neurons have in the recognition of intentions: this in the light of the various criticisms that have emerged, especially in the last ten-twelve years (Cook et al., 2014; Hickok, 2009), on the importance attributed to mirror neurons regarding their role in the aforementioned process, or even more important, in that of understanding what is meant by the expression "understanding actions", which presupposes understanding intentions. Identifying goals and intentions requires a generalization on the perceptual characteristics of the observed actions (Thompson et al., 2019). This is because a goal (such as "to grab") or an intention ("to drink"), can be achieved using different types of grip, and most importantly, the same type of grip can be used to accomplish a large number of different goals and intentions. Since there is no one-to-one correspondence between body part configurations, goals, and intentions (Jacob & Jeannerod, 2005), the same pattern of mirror neuron activation cannot simultaneously represent the action, goal, and intention of the other (Thompson et al., 2019a). Some researchers claim to have found that mirror neurons allow a distinction between different targets (Hafri et al., 2017); however, other evidence has shown that mirror neuron brain areas encode different types of actions based on their perceptual characteristics (Nicholson et al., 2017), suggesting that mirror neuron areas appear to be able to encode the targets of observed actions only when those targets are perceptually distinguishable. Furthermore, generalization about the perceptual characteristics of observed actions appears to occur in conjunction with activity in other, non-motor brain regions that are thought not to contain mirror neurons (Wurm et al., 2016; Wurm & Lingnau, 2015; Spunt & Adolphs, 2014; Spunt, Lieberman, 2013). What therefore seems important to underline is that the main error in the scientific literature on mirror neurons is when it is attributed to them a homuncular-like functioning (Mikulan et al., 2014), as for example in the hypothesis of direct correspondence, which states that an action is understood when its observation causes a resonance in the motor system of the observer (Rizzolatti et al., 2001); this is a case in which the mirror system is given an automatic and mandatory mechanism for understanding (Csibra, 2007). It is therefore possible that mirror neurons alone are not sufficient to explain the encoding of the intentions of others, that is, of the mental states underlying the observed actions. However, there is evidence to support the thesis that mirror neurons are involved in identifying the configuration of body parts when we observe an action (Thompson et al., 2019a). Moreover, it's possible that „the information encoded by mirror neurons is then used by different brain areas in order to identify the mental state underlying an observed action” (Thompson et al., 2019b, p. 110). The most recent approaches to the interpretation of the functioning of the mirror neuron system (MNS) see mirror neurons as part of a system or network that goes beyond the motor cortex and extends to other parts of the brain, including those involved in high-level cognitive processes such as mentalization (Salo et al., 2019). This is also due to the evidence found in laboratory on the increase in connectivity between the areas of mirror neurons and those involved in the processing of others' mental states, when participants are asked to infer the intentions underlying an observed action, with respect to the condition in which they have to judge only how an action is performed (Thompson et al., 2019b; Cole et al., 2019; Libero et al., 2014; Cavallo et al., 2015).

5. Motor simulation and language understanding

As anticipated, motor simulation—a process made possible by neural exploitation—is the reactivation of sensorimotor patterns, detached from their motor functions and exploited in cognitive mechanisms different from those for which they evolved (Borghi & Caruana, 2016). From the perspective of embodied cognition, motor simulation is usually understood as an automatic mechanism and is made possible by mirror neurons. Several studies conclude that it is involved in understanding others' intentions (Binkofski & Buccino, 2006; Gallese, 2007): as said, mirror areas, active both during the execution and during the observation of movements, are not only involved in the recognition of an action, but also in understanding the underlying reasons for the action or its intention.

What could instead be the meeting points between motor simulation, communication, and language? First of all, several studies show how language and action are linked together. One of these has to do with the indexical hypothesis (Glenberg & Robertson, 1999), according to which sentences are understood by creating a simulation of the actions underlying them. In one experiment, Glenberg and Kaschak (2002) created a set in which participants had to judge the meaningfulness of sentences describing the transfer of concrete objects, for example "Andy gave you the pizza/you gave the pizza to Andy", and abstract information such as "Liz told you a story/you told Liz a story": half of the sensible sentences described a transfer to the reader, the other half from the reader to someone else. Participants responded using a box with three buttons held in such a way that the buttons were aligned on the forward/back axis: the sentences were read by holding down the central button with the desired hand. In one condition, the sensible response was made by moving a hand towards the distant button, which then required a movement consisting in simulating a transfer to another person; in the other condition, the response was made by pressing the nearby button, which required a movement similar to a transfer from another person to the reader. As expected, an interaction was found with the time necessary to judge the meaning of a sentence: judgements were faster when the action implied by the sentence matched the action required for the response (approaching or moving away from the body), and this was true for both concrete and abstract transfer sentences. The authors referred to this interaction as the Action-sentence Compatibility Effect (ACE). ACE-type interactions have also been reported from studies employing the use of hypothetical phrases (De Vega, 2008) (for a critique of ACE, see Morey et al., 2021). These results are then confirmed in neuroimaging studies and in the neuropsychological literature, for example, Bak and Hodges (2003) have dealt with how the degeneration of the motor system associated with a motor neuron disorder—in this case referring to amyotrophic lateral sclerosis (ALS)—influences the understanding of action verbs more than nouns; other studies refer instead to the use-induced plasticity of the motor system in influencing the processing of concrete and abstract language (Glenberg et al., 2008), or to the early activation of the motor system following the presentation of a stimulus (Pulvermuller, 2008).

It has been shown that in humans the observation of actions performed with different effectors (hand, foot, mouth) involves the same motor representations that are active during the execution of those same actions (Buccino et al., 2001): this has provided further evidence of the existence of the mirror system in humans, which in our species is not confined only to the Broca area (corresponding to the premotor area F5 of the macaques), but also includes the parietal lobe. Furthermore, an activation of the mirror system is observed, caused by the simple perception of the sound of an action or even when the actions are described verbally (Rizzolatti & Craighero, 2004; Buccino et al., 2004b, 2006); there was also found a somatotopic organization and an overlap between the motor areas activated during the observation of the actions and the motor areas activated during the understanding of the sentences describing those actions (Aziz-Zadeh et al., 2006). These latest studies provide solid evidence for the thesis that motor simulation plays a role in language understanding (we will also see others). The idea is that when individuals listen to words or phrases that imply actions, a modulation of the mirror system should correspond: the effect of this modulation would then influence the excitability of the primary motor cortex and therefore the production of the movements it controls (Buccino et al., 2005; Hauk et al., 2004; Tettamanti et al., 2005).

5.1. Motor simulation and language acquisition

There are some studies and theories in defence of the thesis according to which motor simulation (that is, the resonance mechanism of the motor cortex allowed by mirror neurons) is involved in the process of language acquisition during development. One of these is Gallese’s and Glenberg's (2012) Action-based Language (ABL) theory. In summary, it predicts that it should be easier for infants to learn the names of actions and objects with which they have already learned the appropriate modes of interaction (Huttenlocher et al., 1983). Here we report an example made by the authors (Gallese & Glenberg, 2012) concerning how an infant, who already knows the practical ways of drinking, could learn the verb “to drink”. At the moment when an infant is drinking from a bottle, the parent could say "good drink!": the child's mirror neurons would be activated by the parent's speech act and a hebbian learning process would begin to establish connections between the control of the action aimed at drinking and motor representation of the vocal signal. Then the parent might say “drink (from) your bottle!”: if the child has already learned the name “bottle”, then she may direct her attention to the bottle, grab it and start drinking. Suppose instead that the child focuses on the unknown word "drink" and does not engage in the corresponding action. At this point the parent could say "look, this is what drinking means", and then mimic the act of drinking from the bottle: since the child already knows how to drink, her mirror system would activate the controller necessary for drinking, making possible therefore also in this case a hebbian learning between the modules of the word and those of the action.

As already mentioned, the ABL model for verb learning predicts that infants learn verbs more efficiently if they first learned the corresponding actions: Angrave and Glenberg (2007) found data consistent with this prediction using data taken from MacArthur Child Development Inventory. They estimated the average age, in months, for the acquisition of actions such as drinking, scouting, reading, and the average age of the production of the corresponding verbs: the correlation between the two was very strong, thus making it possible to conclude that the development of word went hand in hand with the development of action (although there was a gap between the production of the action and the production of the word. For an explanation of the reason for this gap, see Angrave & Glenberg, 2007; Gallese & Glenberg, 2012; Wolpert & Kawato, 1998).

The idea is therefore that the sensorimotor system is involved in the perception of action, an activity that would be at the basis of verb acquisition (Pulverman et al., 2006). On the other hand, some studies show how, in adults, the sensorimotor processing of verbs can be a result of associative learning (Cooper et al., 2013; Heyes, 2010), for example as a consequence of training through the use of action pseudoverbs: what is taking place here is the mapping of new verbs onto unfamiliar actions (Fargier et al., 2012). In general, associative learning is involved in the sensorimotor processing of action representations already at an early stage of development, as evidenced by studies showing the role of associative learning in the sensorimotor processing of sounds linked to actions in children between 7 and 9 months (Gerson et al., 2015; Paulus et al., 2013, 2012). Finally, a recent study by Antognini and Daum (2019) showed that toddlers' (18 and 24 months old) sensorimotor system is active during the processing of action-related verbs, concluding that the sensorimotor system plays a role in the processing of action verbs during initial phase of linguistic acquisition. In fact, as pointed out by the authors, the first verbs learned by toddlers are „to a great extent verbs that describe observable actions of people” (p. 82), while the more abstract ones are learned later.

6. Motor simulation: phono-articulatory level, semantic level and ostensive communication. Recognizing intentions in language

So let us now recap some of the statements we have encountered so far:

1) What made ostensive communication possible (i.e. the expression and recognition of communicative and informative intentions) is also what made language possible.

2) Mirror neurons are involved in the recognition of intentions, through the mechanism of motor simulation.

3) Motor simulation is involved in language understanding and in language acquisition.

My conclusion is that the recognition of communicative and informative intentions, through motor simulation, may play an important role in language acquisition, and that it might have played a role in language evolution (see Figure). In the first case (communicative intention) we would have a motor simulation at the phono-articulatory level, occurring with activation of motor cortex areas involved in speech production, which could have a role in the recognition of communicative intentions in verbal communication, a role that I assume to be important for language acquisition during development. What is the evidence for a motor simulation at the phono-articulatory level? In a TMS experiment (Fadiga et al., 2002) it was highlighted how listening to the phonemes induces an increase in motor evoked potentials (MEPs) amplitude registered from muscles of the tongue normally involved in their production: the result was interpreted as a resonance mechanism acoustically connected to the phonological level, a phenomenon confirmed by several other studies (Gallese, 2007). Furthermore, McGuigan and Dollins (1989) showed by electromyography that the muscles of the tongue and lips are activated in the same way both during normal speech production and in covert speech.

What I would like to underline here is that, similarly to what happens in cases of activation of the mirror system and motor areas when listening to sounds linked to actions, phono-articulatory resonance is involved in the processing of the communicative act in itself (Fischer & Zwaan, 2008). „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" (p. 837). The hypothesis is therefore that this mechanism plays a role in the recognition of communicative intentions. However, this does not mean that motor simulation necessarily *always* (i.e. in any case) has a role in the recognition of communicative intentions (or even informative intentions, as we will see soon) during linguistic communication, but that it plays an important role (for the recognition of communicative and informative intentions) for the purpose of language acquisition. When a child listens to a word or utterance, as a consequence there is a resonance at the level of the phono-articulatory system, although the recognition of a communicative intention may occur in different ways—for example through the perception of facial expressions or gestures, even in linguistic communication (Wilson & Sperber, 2002)—my hypothesis is that the communicative resonance mechanism plays an important role in the development, to ensure that attention is directed to language and not to other systems of communication. More specifically, this is the recognition of a *linguistic* communicative intention: it is possible that this mechanism is the basis for the recognition of a linguistic informative intention.

In the second case (informative intention) we would have a semantic simulation, occurring with activation of motor cortex areas involved in processing action content words, which would have a role in the recognition of informative intentions in verbal communication, a role that I assume—also in this case—to be important for language acquisition during development, insofar as words map onto actions. This hypothesis is compatible with the idea that the environment in which our ancestors lived triggered selection pressures in favour of expression of vocal information with action content: communication and language evolved for the purpose of action (Borghi & Caruana, 2016). So, semantic level motor simulation made it possible to think that language understanding has an embodied basis: as regards words and phrases that express action contents, neural structures that preside over the execution of the action could play a role also in understanding the semantic content of the same actions when they are verbally described. As we saw, this emerges from the study by Glenberg and Kaschak (2002), the so-called Action-sentence Compatibility Effect. Studies with TMS showed that pronouncing names of tools, as opposed to those of animals, differentially activates the left middle temporal gyrus, which is also activated with action tasks, as well as the left premotor cortex, which is activated generally when participants imagine themselves grasping objects with their dominant hand (Martin et al., 1996). Other studies instead show that exposure to words that indicate actions or tools produces a motor resonance, which manifests itself with an activation of the motor areas. Exposure to action verbs and words referring to tools semantically related to actions produces a stronger activation of the frontal-central cortical area than exposure to words referring to objects (Martin et al., 1996; Preissl et al., 1995; Pulvermuller et al., 1999). Specifically, action words related to movements of the face, arms or legs (Hauk et al., 2004), activate the fronto-central cortex in a somatotopic way, coherently with the affirmation that the sensorimotor cortex processes aspects of the meaning of words related to action (Pulvermuller, 2005). Further evidence of the automatic activation of motor representations following exposure to action verbs comes from a study conducted with high-density magnetoencephalography (Pulvermuller et al., 2005): subjects were engaged in a task with a distractor as they listened to words that denoted actions involving the leg or face. Different patterns of cortical activation were identified for words referring to leg or face in premotor areas: stimuli for the face-words activated lower front-central areas much more than for the leg-words, while for the opposite an activation of the upper central areas was highlighted. In addition, activations occurred 170 ms after the start of words. Pulvermuller and colleagues (2005) interpreted the results as a reflex of the cortical somatotopic arrangement of motor actions signified by the words. This shows that access to meaning in the recognition of action words is an early automatic process, evidenced by the space-time indications of the activity evoked by the words (Fischer & Zwaan, 2008).

It is important here to highlight that semantic simulation is involved in the recognition of informative intentions since, through associative learning, it is possible to create a correspondence between the recognition of the goal of the action and the word intended to express the content of that action. This is also possible because the ability to perceive actions emerges quite early in development, during a prelinguistic stage (Antognini, Daum, 2019). Furthermore, infants already at the age of 6 months „perceive actions as being directed towards goals". What is important here, therefore, is not to demonstrate the role of motor simulation in every single aspect of linguistic processing (a controversial hypothesis that in recent times has been replaced by dual theories of understanding. See Paternoster & Calzavarini, 2020), but rather its role in the recognition of informative intentions—certainly linked to the semantic processing—of words with action content. In fact, this is important because, as already mentioned, the first verbs learned by infants are largely words that refer to observable actions (Antognini & Daum, 2019).

I would like to underline that my hypothesis is that in both cases, phono-articulatory simulation and semantic simulation, what are simulated are, respectively, the communicative intentions and the informative intentions of verbal communication. Simulation at the phono-articulatory level, as a resonance of the human communicative system linked to the production of speech, could trigger the recognition of communicative intentions, while semantic simulation, being sensitive to the content of the words, could trigger the recognition of informative intentions. It is difficult to say to what extent these mechanisms are involved in language processing, but if expression and recognition of communicative and informative intentions are at the basis of the production/understanding between sender and recipient in language (Scott-Phillips, 2015), then motor simulation must have had a role in the evolution of language, in particular in the transition from manual gestures to vocal gestures. This may have been the initial infrastructure that led to the use of recursive mindreading in ostensive communication (Scott-Phillips, 2015).

7. Conclusions

As I tried to demonstrate in this paper, there are two subsets of motor simulation involved in recognizing informative and communicative intentions: phono-articulatory simulation and semantic simulation. The first, which occurs with activation of motor cortex areas involved in speech production, is involved in the recognition of communicative intentions; the second, which occurs with activation of motor cortex areas involved in processing action content words, has a role in the recognition of informative intentions. The hypothesis is that both have a role in the acquisition of language during development, that is, by means of the recognition of intentions through motor simulation. As already seen, some experiments have tested the hypothesis according to which embodied theories of language comprehension predict that when individuals listen to words or phrases that imply actions, a modulation of the mirror system should correspond (Buccino et al., 2005; Hauk et al., 2004; Tettamanti et al., 2005): in turn this would affect the activation of the primary motor cortex. Overall, several studies support the finding that motor resonance occurs automatically during exposure to words with action content (nouns, verbs, adjectives) (Fischer & Zwaan, 2008).

A goal of future research may be, in addition to testing the role of intention recognition through motor simulation in toddlers' language acquisition, to check the involvement of motor simulation when we infer communicative intentions and informative intentions during verbal communication. One hypothesis is that phono-articulatory simulation is a mechanism, if not sufficient, at least necessary for communicative intentions recognition. As far as semantic simulation is concerned, is it also a necessary mechanism? Both of these aspects of embodied ostensive communication must be tested in the laboratory.

The second hypothesis to emerge from this paper is that, from the point of view of biological evolution, the environment in which our ancestors lived triggered selection pressures in favour of expression of vocal information with action content: communication and language evolved for the purpose of action. This could partly explain the experimental evidence showing the link between motor simulation and words / phrases with action content. Another aspect to deepen may be to understand the role played by motor simulation in the evolution of language, within the framework of embodied ostensive communication.

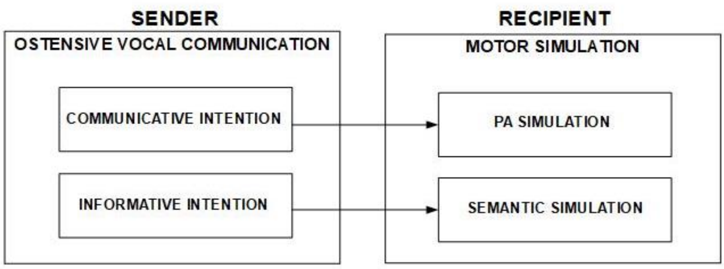


Figure: A diagram that shows how motor simulation may have guided the understanding and evolution of vocal communication.

References

Angrave L.C., Glenberg A.M. (2007). Infant Gestures Predict Verb Production One Year Later. *Paper presented at the annual meeting of the American Psychological Association.*

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

Apperly I.A. (2011). *Mindreaders: the cognitive basis of "theory of mind"*. Psychology Press, New York.

Arbib M.A. (2012). *How the brain got language: The mirror system hypothesis*, New York (NY), Oxford University Press.

Arbib M.A. (2013). Complex imitation and the language-ready brain. *Language and Cognition, 5*, 273–312.

Arbib M.A., Rizzolatti G. (1997). Neural expectations: A possible evolutionary path from manual skills to language. *Communication and Cognition, 29*, 393–424.

Armstrong, D. F., Stokoe,W. C., & Wilcox, S. E. (1995). Gesture and the nature of language. Cambridge, UK: Cambridge University Press.

Aziz-Zadeh L., Wilson S.M., Rizzolatti G., Iacoboni M. (2006). Embodied semantics and the premotor cortex: Congruent representations for visually presented actions and linguistic phrases describing actions. *Current Biology, 16*, 1818-1823.

Bak T.H., Hodges J.R. (2003). The effects of motor neurone disease on language: Further evidence. *Brain and Language, 89*, 354-361.

Bambini V. (2017). *Il cervello pragmatico,* Roma, Carocci.

Bernardis P., Gentilucci M. (2006). Speech and gesture share the same communication system. *Neuropsychologia, 44*, 178–190.

Binkofski F., Buccino G. (2006). The role of ventral premotor cortex in action execution and action understanding. *J. Physiol. Paris 99*, 396–405.

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

Buccino G., Binkofski F., Fink G.R., Fadiga L., Fogassi L., Freund H.J., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience, 13*(2), 400-404.

Buccino G. Binkofski F., Riggio L. (2004b). The mirror neuron system and action recognition. *Brain and Language, 89*(2), 370-376.

Buccino G., Riggio L., Melli G., Binkofski F., Gallese V., Rizzolatti G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cogn. Brain Res. 24*, 355–363.

Buccino G., Solodkin A., Small S. (2006). Functions of the mirror neuron system: Implications for neurorehabilitation. *Cognitive Behavioral Neurology, 19*, 55-63.

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

Cavallo, A., Lungu, O., Becchio, C., Ansuini, C., Rustichini, A., & Fadiga, L. (2015). When gaze opens the channel for communication: Integrative role of IFG and MPFC. *NeuroImage*, *119*, 63–69.

Chomsky N. (1957). *Syntactic structures*, Princeton, Mouton and Co.

Cole, E.J., Barraclough, N.E., Andrews, T.J. (2019). Reduced connectivity between mentalizing and mirror systems in autism spectrum condition. *Neuropsychologia 122*, 88–97.

Cook R., Bird G., Catmur C., Press C., Heyes C. (2014). Mirror neurons: from origin to function. Behav. *Brain Sci. 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.

Corballis, M. C. (2002). From hand to mouth: The origins of language. Princeton, NJ: Princeton University Press.

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*, Oxford, Oxford University Press 2007 (pp. 453–459).

De Vega M. (2008). Levels of embodied meaning. From pointing to counterfactuals. In: De Vega M., Glenberg A.M., Graesser A.C., *Symbols, Embodiment, and Meaning*. Oxford, Oxford University Press 2008 (pp. 285-308).

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.

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

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.

Fogassi L., Ferrari P. F., Gesierich B., Rozzi S., Chersi F., Rizzolatti G. (2005). Parietal lobe: from action organization to intention understanding. *Science, 302*, 662–667.

Gallese V. (2003). A neuroscientific grasp of concepts: from control to representation. *Phil. Trans. R. Soc. B, 358*, 1231–1240.

Gallese V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. Phil. *Trans. R. Soc. B., 362*(1480), 659–669.

Gallese V., Glenberg A.M. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex, 48*(7), 905–922.

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

Gentilucci M., Bernardis P., Crisi G., Volta R. D. (2006). Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. *J. Cogn. Neurosci., 18*, 1059–1074.

Gentilucci M., Santunione P., Roy A. C., Stefanini S. (2004a). Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. *Eur. J. Neurosci. 19*, 190–202.

Gerson, S. A., Bekkering, H., & Hunnius, S. (2015). Short-term motor training, but not observational training, alters neurocognitive mechanisms of action processing in infancy. *Journal of Cognitive Neuroscience*, *27*(6), 1207-1214.

Glenberg A.M., Kaschak M.P. (2002). Grounding language in action. *Psychonomic Bulletin and Review, 9*(3), 558-565.

Glenberg A.M., Robertson D.A. (1999) Indexical understanding of instructions. *Discourse Processes, 28*, 1-26.

Glenberg A.M., Sato M., Cattaneo L. (2008). Use-induced motor plasticity affects the processing of abstract and concrete language. *Current Biology, 18*(7), R290-R291.

Goldin-Meadow S. (1999). The role of gesture in communication and thinking. *Trends Cogn. Sci., 3*, 419–429.

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

Hafri A., Trueswell J.C., Epstein R.A. (2017). Neural representations of observed actions generalize across static and dynamic visual input. *J. Neur*osci. 37, 3056–3071.

Hewes, G.W. (1973). Primate communication and the gestural origins of language. *Current Anthropology 14*, 5–24.

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. *J. Cogn. Neurosci. 21*, 1229–1243.

Huttenlocher J., Smiley P., Charney R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review, 90*(1), 72-93.

Iacoboni M., Molnar-Szakacs I., Gallese V., Buccino G., Mazziotta J., Rizzolatti G. (2005). Grasping the intentions of others with one's owns mirror neuron system. *PloS Biol. 3*, 529–535.

Jacob P., Jeannerod M. (2005). The motor theory of social cognition: A critique. *Trends Cogn. Sci. 9*, 21–25.

Kendon, A. (2011). Some modern considerations for thinking about language evolution: a discussion of the evolution of language by Tecumseh Fitch. *Public J. Semiotics 3*(1), 79–108.

Libero, L.E., Maximo, J.O., Deshpande, H.D., Klinger, L.G., Klinger, M.R., Kana, R.K. (2014). The role of mirroring and mentalizing networks in mediating action intentions in autism. *Mol. Autism 5* (1), 50.

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

Masataka N. (2001). Why early linguistic milestones are delayed in children with Williams syndrome: late onset of hand banging as a possible rate-limiting constraint on the emergence of canonical babbling. *Devel. Sci. 4*, 158–164.

McGuigan F. J., Dollins A. B. (1989). Patterns of covert speech behavior and phonetic coding. *Pavlovian J. Biol. Sci. 24*, 19–26.

McNeill, D. (2012). How Language Began: Gesture and Speech in Human Evolution. Cambridge University. Press, Cambridge.

Mikulan E.P., Reynaldo L., Ibanez A. (2014). Homuncular mirrors: misunderstanding causality in embodied cognition. *Front. Hum. Neurosci., 8*, 1-4.

Morey, R. D., Kaschak, M. P., Díez-Álamo, A. M., Glenberg, A. M., Zwaan, R. A., Lakens, D., ... & Ziv-Crispel, N. (2021). A pre-registered, multi-lab non-replication of the action-sentence compatibility effect (ACE). *Psychonomic bulletin & review*, 1-14.

Nicholson T., Roser M., Bach P. (2017). Understanding the goals of everyday instrumental actions is primarily linked to object, not motor-kinematic, information: Evidence from fMRI. *PLoS One, 12*, 1–21.

Paternoster A., Calzavarini F. (2020). *Comprendere il linguaggio*. Il mulino.

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.

Poizner H., Klima E.S., Bellugi, U. (1987). *What the hands reveal about the brain,* Cambridge MA, MIT Press.

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

Premack D., Woodruff G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526.

Pulverman, R., Hirsh-Pasek, K., Golinkoff, R. M., Pruden, S., & Salkind, S. (2006). Conceptual foundations for verb learning: Celebrating the event. *Action meets word: How children learn verbs*, *2010*, 134-159.

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

Pulvermuller F. (2008). Grounding language in the brain. In: De Vega M., Glenberg A.M., Graesser A.C., *Symbols, Embodiment, and Meaning*. Oxford, Oxford University Press 2008 (pp. 85-116).

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.

Rizzolatti G., Arbib M.A. (1998). Language within our grasp. *Trends in Neurosciences, 21*(5), 188–194.

Rizzolatti G., Craighero L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192.

Rizzolatti G., Fogassi L., Gallese V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat.Rev. Neurosci., 2*, 661–670.

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

Sandler, W. (2013). Vive la différence: sign language and spoken language in language evolution. *Lang. Cogn. 5*(2–3), 189–203.

Scott-Phillips T. (2015). *Speaking our minds*, Palgrave Macmillan (trad. it. *Di' quello che hai in mente*, Roma, Carocci 2017).

Shannon C.E., Weaver W. (1949). *The mathematical theory of communication*, Illinois, The University of Illinois Press.

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

Spunt R.P., Adolphs R. (2014). Validating the why/how contrast for functional MRI studies of theory of mind. *Neuroimage, 99*, 301–311.

Spunt R.P., Lieberman M.D. (2013). The busy social brain: evidence for automaticity and control in the neural systems supporting social cognition and action understanding. *Psychol. Sci., 24*, 80–86.

Stokoe, W. C. (2002). Language in hand: Why sign came before speech. Washington, DC: Gallaudet University Press.

Tettamanti M. et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci. 17*, 273–281.

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 understanding. *Neuroscience & Biobehavioral Reviews*, *105*, 106-114.

Tomasello, M. (2008). The origins of human communication. Cambridge, MA: MIT Press.

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

Blackwell.

Wolpert D.M., Kawato M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks, 11*(7), 1317-1329.

Wurm M.F., Ariani G., Greenlee M.W., Lingnau A. (2016). Decoding concrete and abstract action representations during explicit and implicit conceptual processing. *Cereb. Cortex, 26*, 3390–3401.

Wurm M.F., Lingnau A., (2015). Decoding actions at different levels of abstraction. *J. Neurosci., 35*, 7727–7735.

Zywiczynski, P., Gontier, N., & Wacewicz, S. (2017). The evolution of (proto-) language: Focus on mechanisms.

Angelo D. Delliponti is a graduate in Cognitive Science of Communication and Action, the master's degree course at the Department of Philosophy, Communication, and Entertainment of the Roma Tre University. He is a PhD student in Linguistics at Academia Copernicana, University Centre of Excellence IMSErt in Toruń. The title of his thesis was "Ostensive-inferential communication and embodied cognition". He worked in CosmicLab—a laboratory headed by Prof. Francesco Ferretti in Roma Tre University—on a behavioural experimental design aimed at investigating the comprehension of visuo-spatial metaphoric expressions in adults with visual impairments. Specifically, the expe-

riment intended to evaluate whether a narrative context can facilitate metaphoric comprehension. He partecipated in the THUS 2021—Torun Humanities and Social Sciences Summer Programme—of the Nicolaus Copernicus University on experimental semiotics.

2.2. Motor Simulation and Ostensive-inferential communication: insights and clarifications

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 non-trivial 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.

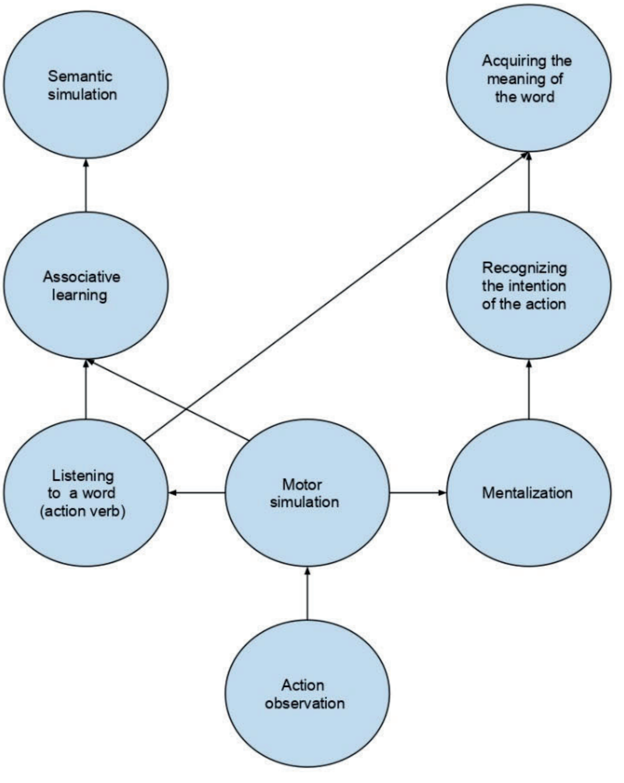


Figure 1. Semantic simulation as a byproduct of recognizing the informative intention in language acquisition

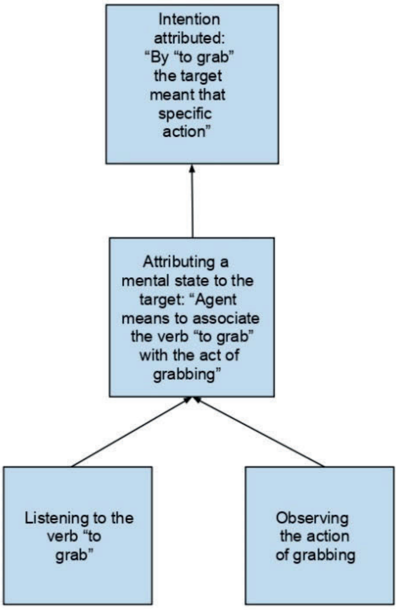


Figure 2. A representation of the full-blown theory of mind

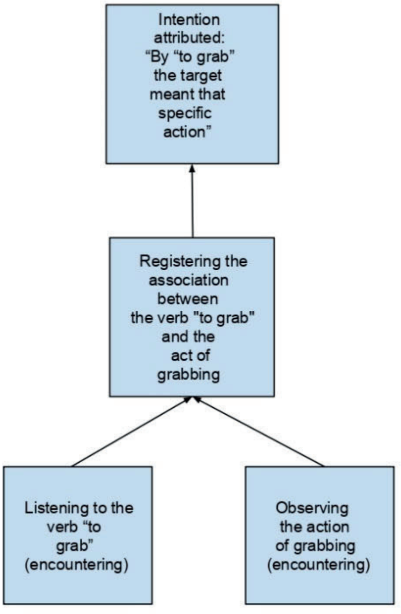


Figure 3. A representation of the minimal theory of mind

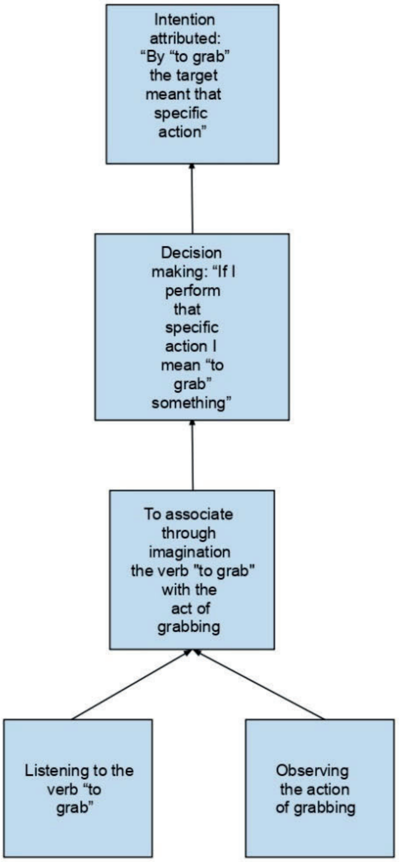


Figure 4. A representation of the high-level simulational mindreading

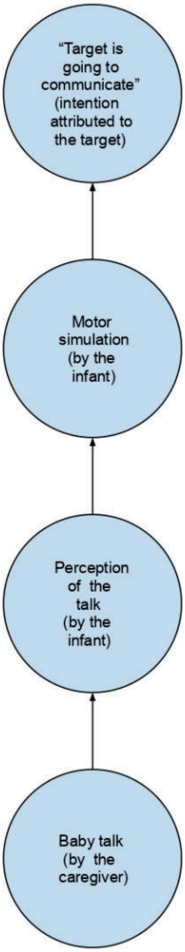


Figure 5. A representation of the low-level simulational mindreading

2.3. Experimental Semiotics: A Systematic Categorization of Experimental Studies on the Bootstrapping of Communication Systems

Biosemiotics
https://doi.org/10.1007/s12304-023-09534-x

REVIEW

Experimental Semiotics: A Systematic Categorization of Experimental Studies on the Bootstrapping of Communication Systems

Angelo Delliponti[1] · Renato Raia[2] · Giulia Sanguedolce[2] · Adam Gutowski[1] · Michael Pleyer[1] · Marta Sibierska[1] · Marek Placiński[1] · Przemysław Żywiczyński[1] · Sławomir Wacewicz[1]

Received: 19 October 2022 / Accepted: 1 May 2023
© The Author(s) 2023

Abstract

Experimental Semiotics (ES) is the study of novel forms of communication that communicators develop in laboratory tasks whose designs prevent them from using language. Thus, ES relates to pragmatics in a "pure," radical sense, capturing the process of creating the relation between signs and their interpreters as biological, psychological, and social agents. Since such a creation of meaning-making from scratch is of central importance to language evolution research, ES has become the most prolific experimental approach in this field of research. In our paper, we report the results of a study on the scope of recent ES and evaluate the ways in which it is relevant to the study of language origins. We coded for multiple levels across 13 dimensions related to the properties of the emergent communication systems or properties of the study designs, such as type of goal (*coordination versus referential*), *modality* of communication, absence or presence of *turn-taking*, or the presence of *vertical vs. horizontal transmission*. We discuss our findings and our classification, focusing on the advantages and limitations of those trends in ES, and in particular their ecological validity in the context of bootstrapping communication and the evolution of language.

Keywords Linguistics · Experimental Semiotics · Language Origins · Communication Systems · Semiotic Game · Language Evolution

Extended author information available on the last page of the article

Published online: 15 May 2023

Introduction

Experimental Semiotics (ES) "focuses on the experimental investigation of novel forms of human communication [...] which people develop when they cannot use pre-established communication systems" (Galantucci & Garrod, 2011, p. 1). A typical semiotic experiment is a game in which participants attempt to communicate without using conventional signs. As a familiar example, think of the popular party game of charades, where the actor (in ES, often technically referred to as the director) must convey without using language a movie title to an audience of guessers (matchers), using to that end non-linguistic resources such as gestures, facial expressions and other bodily movements. Of course, the repertoire of ES games is much broader, involving, for example, the use of drawings, abstract shapes or non-linguistic vocalizations.

A key advantage of ES is that it provides researchers with tools to investigate communication in very general terms, distinguishing the mechanisms of different communication systems (Galantucci et al., 2012a, b). This makes ES relevant for a range of fields in cognitive science and the study of communication and makes its results particularly valuable for several subfields of linguistic research. For example, this approach shows that different factors that operate during communication (such as the structure of meanings or biases for alignment between interlocutors) and during transmission (including population dynamics and constraints on learning) contribute to structures that emerge under different circumstances and shape patterns of variation in languages (Tamariz, 2017).

ES is a successor of an earlier tradition of laboratory studies on language acquisition and change. In a recent review, Nölle and Galantucci (2021) list several 20th century laboratory experiments on communication whose paradigms resemble those used by ES, albeit they have a different focus[1]. The other close intellectual cousin of ES is Experimental Pragmatics (see esp. Galantucci & Garrod, 2011). Experimental Pragmatics investigates how humans use pre-established forms of communication, such as spoken language in dialogic form, with a specific focus on pragmatic aspects of communication like relevance, degree of specificity, and structure. However, the main difference is that ES focuses on the emergence of novel forms of communication so it prevents participants from using natural language, while Experimental Pragmatics studies already existing forms of communication in various contextual situations.

Particularly interesting is the role that Experimental Semiotics has come to play in studies on the evolutionary origins and development of language. At its core, ES investigates the process of creating the relation between signs and their interpreters as biological, psychological, and social agents (see e.g., Morris, 1938 or the notion of "intermediary pragmatics" in Bar-On, 2021). The relationship between symbols and referents, which makes symbolic thought distinct from mere association-making

[1] These include the experiments done by Bartlett (1932) and Bavelas (1950, 1952), who studied how information transmission leads to the creation of 'stereotypical' memory patterns or the emergence of leadership roles, and the early attempts of using artificial miniature languages to study language change (e.g., Esper 1925; Wolfle, 1933).

processes, was seen as the hallmark of human reasoning and language, and essential for their phylogenetic development. It has been of interest since ancient times, with the key question being whether it was an innate and universal ability or the outcome of complex social processes. Several unverified stories speak to this interest in recreating the process of how a communication system is born: namely through children being raised in linguistic isolation without any human interaction (see e.g., Żywiczyński, 2018). Such a cruel procedure, known as the "forbidden experiment", was allegedly performed by Psamtik I of Egypt, Frederick II (Hohenstaufen) of Sicily, and James IV of Scotland (Campbell & Grieve, 1981). The only conclusion to be drawn is that language is not entirely innate: if a child is completely deprived of any linguistic input, they will not speak any language (Galantucci, 2017). In a way, ES could be seen as a descendant of these stories about the "forbidden experiment" that is, however, following modern-day ethical standards. Its goals consist in studying how communication systems are brought into existence and how sets of conventional relations are created and then shaped through repeated use, which in turn helps to understand the processes that underlie the development of language in ontogeny, cultural-historical change and phylogeny (Galantucci, 2005).

Over the years, ES has grown into an extensive field and, as such, has become the object of several overviews (e.g., Galantucci et al., 2012a, b; Galantucci & Roberts, 2012; Galantucci 2017). In an early summary of ES research, Galantucci et al. (2012a, b) outlined the basic research problems tackled in the field, described the main study paradigms, and explained their implications for linguistics. They described three types of experimental paradigms: semiotic referential games, semiotic coordination games, and semiotic matching games. The authors also recognised five main research themes of ES: "the emergence of linguistic structure, the role of interaction in communication, the role of inter- and intragenerational processes in the evolution of language, the study of sociolinguistic processes in the laboratory, and the bootstrapping of communication" (Galantucci et al., 2012a, b, p. 581). To demonstrate the potential of ES as a major complement to linguistic research, Galantucci and colleagueGalantucci et al. (2012a, b) specified three reasons: enabling the study of novel communication systems; providing full access to the history of their development; and the potential for easily controlling the conditions of this development. Another, similar overview of ES recognises three main themes: linguistic properties as the consequence of communication, social factors in communication, and the bootstrapping of communication (Galantucci & Roberts, 2012).

Importantly from the perspective of this paper, all the existing overviews of ES follow the traditional format of the review paper, without applying the tools of a systematic literature review (SLR). Although these existing overviews describe particular paradigms in considerable detail, their narratives and coverage of the literature are necessarily subjective and selective. In this paper, we propose a different, bottom-up, approach to characterizing ES, inspired by the systematic literature review approach (Xiao & Watson, 2019). Our main goal is to create a comprehensive resource of ES studies relevant to the earliest stages of establishing a communication system, which is categorised by a broad range of design parameters. In doing so we aim to create a resource that will inform future ES works, but also to understand how to conduct research and which paradigms have been ignored. Therefore, we conducted a systematic review of 60 ES studies, published over the period of 20 years, from 2002 to 2021. We coded the papers for multiple levels across dimensions related to the properties of the emergent communication system and the properties of the study design. The results of the coding were subjected to statistical data analyses as categorical variables. Other dimensions were textual and reflected the more general and qualitative aspects of the papers, such as their main findings. All the coding was compiled into a single, interoperable and reusable dataset, as is described in detail below. Thanks to these efforts, we are able to provide a novel, systematic approach to characterize the properties of ES studies.

Dataset

Inclusion Criteria, Acquisition of Articles, Coding Procedure

Inclusion Criteria

Although a classic understanding of ES restricts its meaning to "controlled studies in which human adults develop novel communication systems" (Galantucci et al., 2012a, b, p. 477), this definition is occasionally extended onto controlled studies in which adults "impose novel structure on systems provided to them" (Galantucci et al., 2012a, b, p. 477). On this broader definition, ES also subsumes studies where sign-meaning pairings are already provided by the experimenters rather than emerging naturally in the game, as in most "alien language" studies using the iterated learning paradigm (e.g., Cuskley, 2019). In line with our interest in language origins, particularly the early bootstrapping phase of communication, we adopted this first – classic and narrower – definition of ES as an inclusion criterion. That is, we included studies on the emergence of novel communication systems and excluded studies in which participants began learning the meanings already assigned by fiat to a set of signs. For practical reasons, we limited articles to those that had been published in peer-reviewed journals, thus excluding experiments reported in chapters in edited volumes or proceedings papers.

Acquisition of Articles

Articles that matched the inclusion criteria were identified and acquired through a three-step procedure. First, an initial list of ES studies consistent with our criteria was compiled bottom-up. In the second step, the coders went through the references of the articles in the initial list, as well as references in review articles ((Galantucci et al., 2012a, b; Galantucci & Roberts, 2012; Galantucci, 2017; Galantucci & Garrod, 2011; Nölle & Galantucci, 2021) to identify articles containing further studies eligible for inclusion. Finally, the coders did a series of targeted searches on Google Scholar and Connected Papers for keywords such as "experimental semiotics," "semiotic game," or "laboratory languages," in order to extend the search to all studies linked by similar topics. The completeness of the list created in steps one through three was later approved by a leading expert in ES external to the coding team.

Coding Procedure

The coders were first trained to apply the coding dimensions to an initial set of two papers each. Their coding was then discussed and refined by all researchers. All papers on the final list were distributed among eight coders. First, they worked independently, each coding the assigned papers for the 13 dimensions described below (Sect. 2.2), and marking potentially difficult classificatory decisions. These were then resolved consensually through discussing such unclear cases in the coding in a group.

Coding Dimensions

The papers were coded for three types of dimensions: (1) basic bibliographic and scientometric information (the year of publication, the total number of citations on Google Scholar as of April 22, 2022, as well as citations per year), which gave us an idea of the popularity of each paper in the field; (2) general information: the paper's main themes or topics, a brief summary of the main findings, the number of participants, their age range, and the experimental setting (laboratory or online); (3) study design properties, which were treated as categorical variables, coded as numerical values assigned to category labels. For example, the variable "type of game" had two values, "1" for referential games and "2" for coordination games. These values were then statistically analysed. The dimensions included in (3) are described in detail below. The coding dimensions are based on descriptions of ES paradigms in the literature as well as the key differences evident between the studies that can be related to overarching type differences.

Type of Games: Referential vs. Coordination

Despite its recent origin, ES has developed two main paradigms: *referential games* and *coordination games*. The referential framework of ES is derived from standard referential communication tasks that were employed in Experimental Pragmatics (see e.g., Krauss & Weinheimer, 1966), in which participants had to converse about novel shapes using natural language. In the ES version, the use of natural language is forbidden, so players must communicate about a predetermined stimulus (e.g., a piece of music or a concept) using other means. In standard referential games, the set of signals used for communication is open, whereas the set of referents to communicate about is closed and pre-established by the experimenters (Galantucci et al., 2012a, b) (see Sect. 2.2.3). The purpose of the communicative act is communication itself; the goal of the director is to have the matcher correctly guess the intended meaning. A paradigmatic model of referential games is the "Pictionary" set-up employed by Garrod and colleagues (2007), in which the director has to graphically depict various concepts and communicate them to the matcher(s).

In coordination games, the communicative act is instrumental for the purpose of the game, which is succeeding in a specific task that usually involves moving an agent in a virtual space and coordinating the moves with the partner. In these games, successful communication can be supported by different sets of referents, therefore players must agree not only on the set of signals but also on the set of referents used to make communication successful ((Galantucci et al., 2012a, b). One model of coordination games has been dubbed the “tacit communication game” (TCG; Galantucci 2017): in TCG, each player of a dyad controls one virtual agent (a geometric figure, the “token,” which can be moved and rotated) over a 3×3 grid, and their goal is to place their tokens in the correct positions, established by the experimenters. Only one of the players, the sender, knows the correct position and has to communicate it to the other using only moves over the board. The moves of the sender thus serve a double function of, first, moving the player token into the correct position and, second, communicating to the other player their correct position. The sender has to find a way to clarify which moves have just an instrumental purpose and which have a communicative purpose.

Vertical Transmission

In most ES studies, a communication task is performed within a dyad or a larger group of participants whose composition remains constant throughout the ES game. However, there is an interesting minority of studies with a dynamic group composition, such that some players leave, and others join the group within the timeframe of the game. Such a design enables the *vertical transmission* of information, which occurs when the communicative output of one generation (e.g., a set of signs they have converged on) becomes the input to which the next generation is exposed. One example are replacement microsociety studies (e.g., Caldwell & Smith, 2012), where the interacting group is composed of a director and a small number of receivers; at the end of each turn, the director is removed from the game and the most experienced matcher becomes the new director, while a new player enters the group as the least experienced matcher. These studies simulate a natural aspect of human society: the communicative conventions created at a given time are passed onto the next generations, which have to learn and inevitably modify them. Inserting the vertical transmission of established conventions into ES designs hence offers a way to study the cultural evolution of sign systems. However, studies that do not feature vertical transmission focus on the emergence of novel communication systems in the interaction of agents engaged in a particular activity, either reference or coordination (e.g., Galantucci et al., 2012a, b).

Signals and Referents

Two dimensions in our coding scheme concern the type of signals adopted in each experiment. The first is about the kind of *medium* employed in communication. A large majority of experiments use either vocalizations, bodily-visual signals (i.e., communicative bodily movements, such as gesture, pantomime, facial expression or gaze), or graphical signals (drawing, symbols, lines, colours, etc.). There was also a small minority of studies whose medium of communication did not fall under any of these three possibilities (e.g., Iizuka et al., 2013).

The category *signal space* was further subdivided into discrete and continuous. In a *discrete signal space*, senders chose the signal from a set of specific, predefined possibilities, often limited in number, effectively making signal production a multiple alternative forced choice task. An example of a discrete signal space is to choose an Arabic numeral from a set of 1 through 10 as the signal to be sent to the receiver. TCGs are games that usually employ a discrete signal space: the possible configurations of the tokens which the sender has to use for communicative purposes are inherently limited because the token can only move on a small grid (e.g., Blokpoel et al., 2012). Conversely, in a *continuous signal space*, senders could produce any signal form possible within the constraints of the communication medium; an example would be pen-and-paper (or digital) drawings, which are not limited to a number of distinct variants but instead can take on any shape. An interesting but much less frequent possibility is that the signal space is continuous but not unlimited; in this case, the director must choose within a spectrum of possibilities, for example, shades of colour (e.g., Roberts & Clark, 2020).

We also coded for what we dubbed the *meaning space* and identified the types of referents used for communication. The referents can be common concepts (objects or actions which are easily verbalizable, like "house," "dog" or "giving a kiss") or more abstract entities (unfamiliar geometric shapes, pieces of music, configurations). We decided to create another level for this category, which mostly applies to coordination games, that is when the referents are a particular position or disposition of the tokens. In Zlatev and colleagues (2017), we have an example of a study with a referential game in which referents made up of meaningful concepts are taken as meaning space. In this case, pantomime was used to express concepts such as a father kissing his daughter, a person hugging another person, etc. On the other hand, in Stevens and Roberts (2019), we have a coordination game, as the sender and the receiver had to coordinate in order to find the best way to communicate and interpret the expressed signs. The meaning space was composed of lines inside the cells; therefore, it was the position of the signs that was communicated. In this sense, we claim that the meaning space refers to a location.

Interaction

We also examined the parameters of interactions between players and the general setup of the game. One of the categories we used for this was related to the *feedback*, which is information about the outcome of the communicative interaction process. As a simple example, if the director produces a clenched-fist gesture, to which the matcher responds "war," in most studies this will be followed by feedback in the form of "correct" (if "war" was indeed the intended meaning) or "wrong" (if the intended meaning was something else). We were interested in the source of this information: in some experiments, feedback comes from the other player(s), in others from the experimenter themselves, and in still others there is no feedback. Sometimes (as in the Pictionary-like game in Fay et al., 2017), the presence of feedback is itself one of the studied variables, as its presence or absence can alter communicative success and other important properties of the exchange.

Related to feedback is the category of *turn-taking*, which describes the turn-order of the players' actions. As one option, there could be no turn structure, with players being free to take their actions at any time and in any order, even simultaneously. However, there could also be a fixed turn structure governing the exchange of turns; a frequent pattern is the director acting first by sending the signal, and then the matcher selecting a possible referent. Such a structure could either be pre-specified, that is, imposed by the mechanics of the experiment or may emerge spontaneously during the game, despite not being formally determined by the experimenters. A different category, *interchangeability*, captured whether the roles of directors and matchers were assigned to particular players for the entire duration of the game, or if players could change their roles. That is, if a player could be the director at one point of the game and then change their role to that of the matcher at another point, interchangeability was present. Conversely, if one player was always the director and the other was always the matcher, interchangeability was absent.

Two further categories are related to the interacting group. *Group size* was the total number of players in a group, whereas *communication type* referred to how many players took part in an individual interaction act. For example, if a study had groups consisting of seven players but each communicative act always happened between two players, *group size* and *communication type* would be classified respectively as "larger groups" and "dyadic communication."

Finally, we examined whether the interaction between senders and receivers in the experimental setup was, or not, *simultaneous*. A simultaneous interaction is when the reception of the signal occurs immediately after its creation by the director, which is a characteristic of live interaction. If the matcher is looking at a stimulus recorded at an earlier time, the interaction is considered non-simultaneous.

Alignment of Interest

The category of alignment of interest was introduced to study one of our greatest conceptual interests: whether the origins of language were marked by a competitive or cooperative use of our communicative means (e.g. Tomasello, 2008; Scott-Phillips, 2014; Ferretti, 2022). Language is traditionally believed to be born out of a cooperative attitude among humans: after all, if signals were used mostly for deceptive purposes, no one would have reason to trust them and language would become useless and disappear. Note that models of animal communication, inspired by Krebs and Dawkins (1984; also Dawkins & Krebs, 1978), mostly see communication as a means to influence and manipulate the behavior of others to one's own advantage: the cooperative presupposition would, in fact, imply an evolutionarily unlikely altruism by the senders of signals or a similarly unlikely gullibility by receivers. Some current models of language function and evolution consider it to be characterized by a mixture of competitiveness and cooperativeness (Sperber et al., 2010; Lee & Pinker, 2010). ES is a particularly well-suited means for studying the emergence of the early properties of human communication, such as compositionality and combinatoriality, under the influence of humans' pragmatic abilities. It would be interesting to know if this development can also occur when there are differing interests among the people involved in communication. An important distinction must be made here: one thing is competition *among* interacting groups (which is sometimes employed as an incentive for players; the group with higher communication success receives more points – sometimes associated with a monetary prize); another is competition *inside* each interacting group, that is, the existence of a conflict of interest between senders and receivers. This latter kind of competition is the one we are interested in as it allows us to study the competitive or cooperative nature of early communication.

Summary

This is a short summary of the dimensions included in the statistical analysis. The numbers associated with each value (which are followed by an explanation of those values) are the actual numbers that were subjected to a cluster analysis. In all cases, a further value "other" was added for papers that did not fit into any of the pre-established levels. In these cases, additional specifications were included (Table 1).

To follow up on our first example, according to these coding dimensions, a typical game of charades would be classified as referential (goal: to be understood, to convey a concept), involving no vertical transmission (or only marginally so, if successive players adopted some gestures and pantomimes used by the previous players, see e.g. Christiansen & Chater, 2022), a bodily-visual medium of communication, an open and continuous signal space (no predefined set of gestures – any bodily configuration can be used), a meaning space of meaningful concepts (such as movie titles), feedback that comes from the director, no turn taking (directors and matchers do not need to wait their turn, can send signals / provide responses at any time and in any order), the presence of interchangeability (people change roles of doing the pantomimes and guessing their meaning), group size and communication type that depend on the number of matchers in the audience; interaction that is simultaneous (unless the pantomimes are recorded and later shown to the matchers), and the interests that are aligned (the director wants the matchers to guess correctly, and so do the matchers).

Applications

The database presented in the above sections is intended as a multipurpose resource with a broad spectrum of diverse applications in Experimental Semiotics research. Here, we limit ourselves to pointing to three avenues in which this resource can be put to use.

Informing Reviews and Designs

Firstly, a basic application of the database is in informing literature reviews on the field – both those intended to provide theoretical overviews and those underlying experimental studies – to facilitate a more systematic and comprehensive coverage of the relevant literature. For example, researchers planning to address their research question through a *coordination game* design will be in a position to instantly identify and access an exhaustive set of previous studies using this particular paradigm. Furthermore, the proposed classifications may help scaffold new experiments in ways that facilitate rigor and productivity. Since the development of experimental designs involves many decisions that are usually taken implicitly, the dimensions used in our database may serve as a guide to reviewing such decisions in an informed manner. For example, planning the design of the said novel coordination game study involves

Table 1 Coding dimensions used in the study

Dimension	Description	Values
Game type	If the goal of the game is to be understood (referential) or if communication is a means of achieving something else (coordination).	1. Referential 2. Coordination
Vertical Transmission	If the study incorporates replacing "generations" of players.	1. Vertical transmission is present 2. There is no vertical transmission
Medium of communication	The medium of the signals.	1. Vocalizations 2. Bodily-visual 3. Graphical
Signal Space	Properties of the set of usable signals as means of communication.	1. Limited and Discrete 2. Open and Continuous 3. Continuous but limited
Meaning Space	Type of meaning of the referents.	1. Meaningful concepts / words 2. Abstract shapes / symbols / configurations 3. Location
Feedback	Source from which participants receive information about the outcome of the game.	1. No feedback 2. Feedback comes from other participants 3. Feedback comes from the experimenters or emerges from the experimental setup itself 4. Feedback is provided by both the system and the participants 5. Other
Turn-taking	Does the study include roles of directors and matchers or not?	1. The roles are pre-established by the experimenters 2. The roles are absent in the study 3. The roles are not formally established by the experimenters, but they emerge spontaneously
Interchangeability	Do the players exchange their roles as directors and matchers?	1. Yes, players alternate between the roles 2. No, players remain in their roles for all the length of the experiment
Group size	Number of people in each interacting group.	1. Pairs 2. Small groups (three to five people) 3. Large groups (more than five people)
Communication type	Number of people involved in each interaction.	1. Dyadic 2. Triadic 3. Four-way
Simultaneity	Is the interaction simultaneous or is the matcher watching a recording?	1. Simultaneous 2. Non simultaneous
Alignment of interest	Is there a conflict of interest inside the interacting group's members?	1. Common interests 2. Conflict of interest

deciding on the medium of the communication, community size, openness of the signal space, and so forth, which can be readily compared against such decisions in existing studies.

It is worth noting that these points generalise beyond the area of Experimental Semiotics, extending into studies on the origins of communication more broadly. This is particularly relevant to agent-based modelling, where building the model itself involves taking explicit decisions on several dimensions, such as signal space, turn-taking, or alignment of interests (e.g. Zubek et al., 2023).

Identifying Patterns in ES Research

Another application with direct implications for research practice is searching the multidimensional space of possible design configurations to identify over- as well as underrepresented designs. These can be either choices in a single dimension or, more interestingly, choices along two or more dimensions that are highly correlated with one another (e.g. *game type: coordination* almost invariably involves *medium of representation: graphical*). By extension, this also allows us to point at alternative design configurations – i.e. ones that are possible in principle but not actually implemented in existing studies – thus showing us unexplored or underexplored possibilities.

As a simple example, consider the dimension *alignment of interests*. Almost all studies conform to the default "cooperative" setting of making the interests of the communicators aligned with each other: all parties of the communicative situation share the same goal of converging on the same referents or locations. There are only two exceptions (dos Santos et al., 2012; Inoue & Morita, 2021), which introduce some degree of conflict of interest (thus, rivalry) between the communicators, who are incentivised to pursue one's own communicative goals even when this might be at the expense of their partners. This seems to be consistent with the theory: that it is difficult to imagine the bootstrapping of a communication system without cooperation. Of the two studies considered, at least in one case, competition had a positive effect on the consolidation of communication. However, this happened when the competition was on a global scale, and not on a local one: the result is that "humans change their level of cooperation as a function of the scale of competition (…), highlighting the importance of considering the scale of competition in studies of cooperation and communication" (dos Santos et al., 2012). Thus, a question could be posed on whether we need more studies with some kind of conflict of interest to investigate the possible role of competition in communication, which would be in line with some recent theoretical proposals on the role that persuasion may have played in the evolution of language (Ferretti & Adornetti, 2021).

To provide a more complex example, we conducted an analysis of correlations between our dimension values. To this end, the data frame was transformed in such a way that each factor level could be encoded as either 0 (= not present in the study) or 1 (= present in the study). For instance, the category "game type" was divided into "referential" and "coordination", each of which were subsequently marked with 0 or 1 s. Several more technical variables, such as the year of publication or population age, were not included in the analysis; others were removed due to being a single-level variable (e.g. "alignment of interests"). Here, we present only a sample of our analysis with the strongest correlations between the encoded categories, and the whole data frame of correlations can be accessed under https://osf.io/ad7b4/?view_only=0590ad2c505840dd8ccebd1d8f890cb4 (Fig. 1).

Our analysis suggests a strong correlation (r = 0.85) between meaning_space_3, i.e. communicating about a location, and Coordination, i.e. studies that investigate

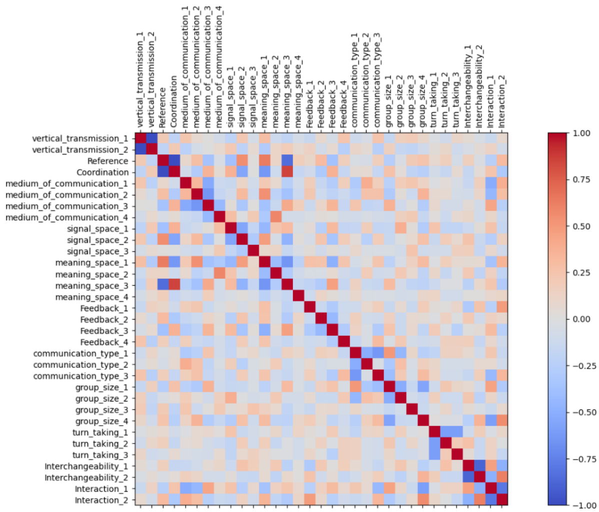


Fig. 1 Correlation matrix of features, where values closer to 1 suggest a positive correlation between two variables, whereas values closer to -1 a negative one

communicative coordination between participants. Referential games, on the other hand, have a strong negative correlation (r = -0.85) with this type of meaning space, and are more frequently used (r = 0.63) in studies where participants have to communicate about meaningful concepts (meaning_space_1). Another strong correlation (r = 0.61) exists between interchangeability_2 (i.e. experiments in which participants switch between roles) and interaction_2 (i.e. communication in dyads).

These results inform us about the limits of particular study designs and those of communication itself. Communicating about location is a complex process that typically requires coordination between participants, whereby they incrementally update their state of knowledge. It would be rather difficult to communicate about the location of an object in a referential game; doing so would perhaps be possible but would require an innovative design. The other strong correlation – between dyadic communication and interchangeability – can reflect a concern for interpreting the results of the study and removing factors related to the number of interlocutors involved in a conversation.

Agglomerative Hierarchical Clustering

Agglomerative hierarchical clustering is an approach adopted in the exploratory analysis of multi-dimensional data (Nielsen, 2016). This method involves building a hierarchical tree from "leaves" - the most basic units - and iteratively builds a hierarchical structure. The leaves of the tree are merged on the basis of the smallest distance between them, then those merged leaves are aggregated into bigger units until the root of the tree is reached (Manning et al., 2008). The output of agglomerative hierarchical clustering is usually depicted in the form of a dendrogram; the dendrogram resulting from the analysis of our dataset can be accessed here: https://cles.umk.pl/evolang-network/dendrogram/.

The dendrogram analysis resulted in a tree structure that can be divided into three broad categories. The first division between papers occurs between a single paper

(Perlman & Lupyan, 2018) and the rest of the database. This divide occurs due to the fact that the study reported in that paper involved relatively rare design choices along several dimensions, such as feedback, group type and turn-taking. The second category in the dendrogram occurs between a sub-branch represented by such studies as Raviv et al. (2019), Garrod et al. (2007), Selten and Warglien (2007). What these studies have in common is their medium of communication (primarily drawings), referential game type, alignment of interests and the rigid assignment of roles (e.g. director and guesser). The last major subgroup consists of such studies as Żywiczyński et al. (2021), and Motamedi et al. (2018; 2019). In this group, a majority of studies were conducted with the use of referential games and a non-verbal bodily medium of communication (i.e. gestures or full-body pantomime).

Experimental Semiotics over the Years

Finally, a meta-level application of the database is in identifying trends across time, to help achieve a deeper understanding of the historical development of Experimental Semiotics. To properly understand the evolution of ES studies, we have conducted a cluster analysis based on 11 features described in Sect. 2: "Presence vs. absence of Vertical transmission?", "Referential vs. Coordination", "Medium of communication," "Signal space," "Meaning space," "Feedback," "Communication type," "Group size," "Participants of the main study: Age," "Turn-taking," "Interchangeability of the signaller/receiver roles". Using the Python programming language (van Rossum & Drake 1995) and the Scikit-learn library (Pedregosa et al., 2011), we ran a *k*-means clustering algorithm that classified papers into six clusters based on these features. The optimal number of clusters was determined using the Elbow Method. Figure 2 shows how the clusters are distributed over the considered decades, with each color corresponding to a specific cluster[2].

One thing that stood out when looking at the coded dimensions is the presence of static categories, that is, those dimensions that varied little or very little in most of the studies analyzed throughout the entire period of time considered. Among the static dimensions, we have "Communication type," which was almost exclusively dyadic (although in recent years, there has been a greater presence of non-dyadic communication); "Presence or absence of Vertical transmission?," which reports only four

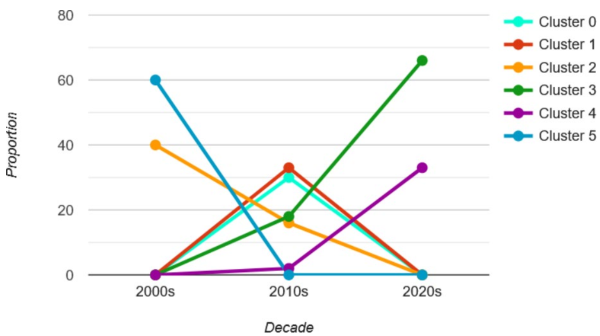


Fig. 2 A graphic representation of the result of clustering, with the x-axis representing the decades, and the y-axis the predominance (in proportion) of a cluster in the considered decade

[2] All visualisations were produced using RapidTables (n.d.).

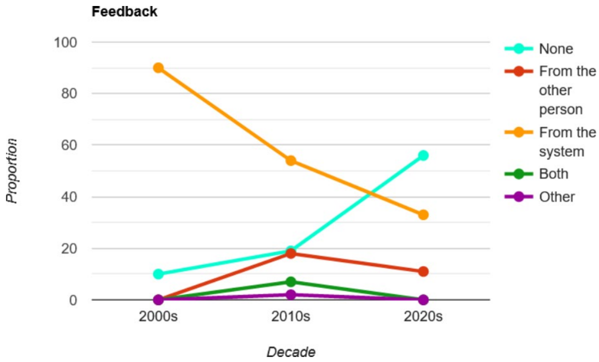


Fig. 3 The "Feedback" dimension (over considered decades)

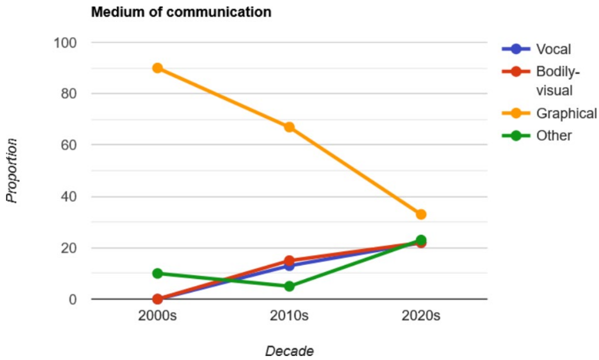


Fig. 4 The "Medium of communication" dimension (over considered decades)

studies containing some transmission of the result of the communication task from one group to another; "Lab or online," with only two studies that were conducted online; "Simultaneous interaction," of which only six studies were not characterized by a contextual interaction; and "Alignment of interests," with only two studies that included conflict of interest.

Regarding dynamic categories, on the other hand, an example is found in the "Feedback" dimension, which for the 2002–2007 period was characterized by a predominance of information received from the experimenter, or more generally, from the system. Only one of these studies reported different values. In Fig. 3, this phenomenon can be observed, with a progressive decrease in the paradigm over the considered decades.

Another salient dimension is "Medium of communication". In the period between 2009 and 2014 this dimension was characterized by the fact that it was made up almost exclusively of studies that used a graphical medium (except for one). In Fig. 4, it is possible to note that there has been a decrease in the number of studies that used a graphical type of communication over the decades.

Two closely related dimensions are "Referential vs. coordination" and "Meaning space." Indeed, starting from 2014, it is possible to observe an almost exclusive use of referential games, which corresponds to an equally preponderant use of meaningful words/concepts of the relative "Meaning space" dimension (See Figs. 5 and 6).

Fig. 5 The "Referential vs. coordination" dimension (over considered decades)

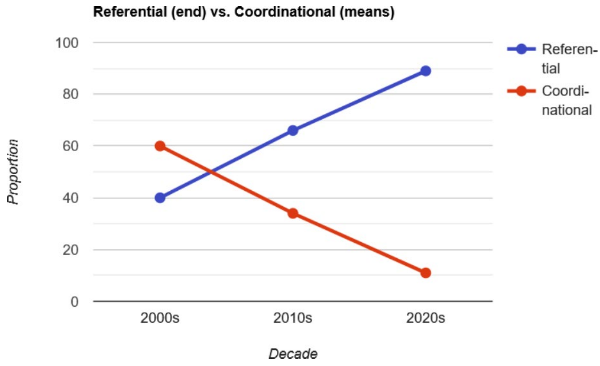
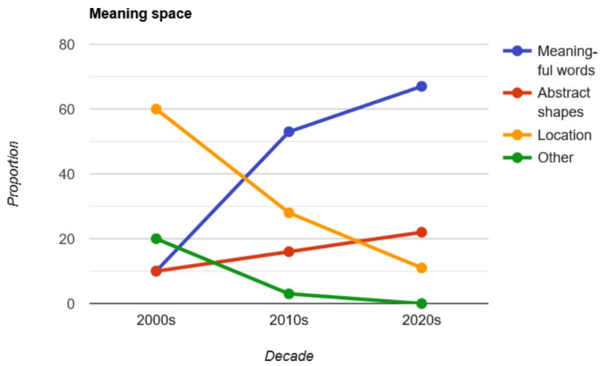


Fig. 6 The "Meaning space" dimension (over considered decades)



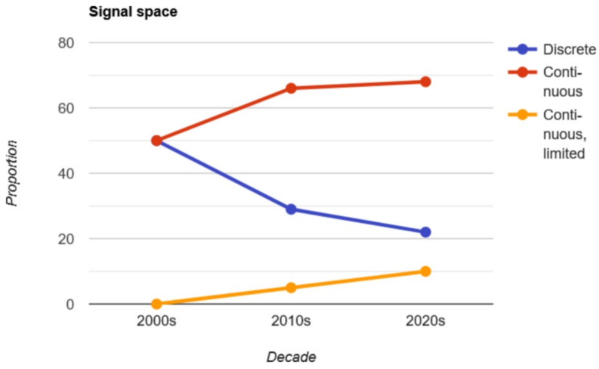
In both categories, it is possible to observe a progressively predominant use of the aforementioned values. This is directly related to the goal of the game: in referential tasks, where the goal of the game is to be understood, it is easier for the relative meaning to be made up of meaningful concepts (or, at most, abstract shapes); in coordination games, where communication is only a means for accomplishing the goal of the game, it is easier to have location as a meaning space.

Another interesting dimension is "Signal space", which starting from 2014 becomes almost exclusively open and continuous (in only one paper this is not the case, (Fig. 7).

ES over the Years: Discussion

One of the research questions we mentioned was whether there is any reason for the presence of static categories, that is, why most ES studies show largely identical values for specific dimensions. While the reasons for the lack of diversity in research designs under certain dimensions can be investigated in more detail in future studies, we offer some preliminary answers. It seems intuitive that dyadic communication is more suitable for the observation of communicative interaction according to the classic sender/receiver model, despite the recent increase in interest in non-dyadic studies. The relative absence of vertical transmission may be in large part due to our inclusion criteria being limited to studies on the creation of new communicative sys-

Fig. 7 The "Signal space" dimension (over considered decades)



tems "from scratch," while vertical transmission is typically studied in the lab with artificial language designs, where the initial signal-meaning pairings are given to the participants (e.g., Kirby et al., 2008). The paucity of studies carried out online could also be because the laboratory is perhaps more suitable for building experimental settings that are ecologically realistic. A similar argument may be valid for studies that do not involve simultaneous interactions.

It would be interesting to explore why there has recently been a drop in the number of studies that make use of the coordination task paradigm. One answer could be that it is more logistically challenging. However, studies that use coordination games are potentially of considerable interest for ES and, more generally for the analysis of the bootstrapping of communication systems, as some kind of alignment of interests is necessary to achieve coordination. Results of such studies could potentially highlight some of the social and cognitive dynamics that underlie communication and language. One potential reason for coordination tasks to be used less frequently is that it is more difficult to establish novel form-meaning pairings for purposes of coordination than for the sole purpose of identifying referents. This is because for referential communication games the potential meaning space is generally prespecified and limited, as opposed to the meaning space required for coordination games which is potentially open-ended.

Starting from 2014, the *signal space* dimension became almost exclusively open and continuous (Fig. 6.) This could be explained by the fact that ES studies seem to become increasingly ecologically realistic over time. The use of open and continuous signals is consistent with important threads in the literature on the evolution of language, for example, related to iconicity and holistic nature of early signs (Perlman et al., 2015). In a study by Nölle and colleagues (2018) gestural communication was used in order to express meanings represented by drawings of characters who belonged to categories delimited by, among other things, shared colours. This is an example of an open and continuous signal space, which was also the case with Zlatev and colleagues (2017).

Some of the research questions we asked relate to the problem of how ES studies have evolved over time: if there are particular trends in specific periods over the examined decades (although experimental semiotics is a rather new research field); if the categories can be related to each other in some way; if there is an explanation we can provide for the observed trends; or why some paradigms are systematically ignored at the expense of others (e.g., why there are so few studies with vertical transmission). Other questions could refer to the results of the studies analysed, for example, whether similar results correspond to similar experimental paradigms or if there is evidence for specific empirical results that correspond to a coherent global picture, whether they are in line with the theoretical proposals, and so on.

Conclusion

In this paper, we present a novel, systematic approach to the characterisation of ES studies according to the coding dimensions of the type of communication game, the presence of vertical transmission, the properties of the signaling and meaning spaces, the type of interaction, and the presence of the alignment of interest. This resulted in a dataset of 60 studies that were coded for these dimensions. In an exploratory analysis, we showed several potential applications of this dataset, including demonstrating how it can be used to examine changes in ES through a cluster analysis of the distribution of coding dimensions over time. This approach, along with the generated annotated dataset, has several potential applications. For example, it allows for a more fine-grained analysis of similarities and differences in the development of novel communication systems depending on the design features of ES studies. It also allows us to measure which dimensions cluster to provide more information about which experimental design is best suited for investigating particular research questions. Overall, an approach that systematically compares the underlying design properties of ES studies can help to specify the different mechanisms that influence the properties of novel, emerging communication systems.

Author Contributions A.D. contributed to conceiving and designing the analysis, collecting and coding the data, conceptualizing and writing the manuscript, and doing the cluster analysis. R.R. contributed to conceiving and designing the analysis, collecting and coding the data, and writing the manuscript. G.S. contributed to conceiving and designing the analysis, collecting and coding the data, and conceptualizing and writing the manuscript. A.G. contributed to conceiving and designing the analysis, collecting and coding the data, and conceptualizing and writing the manuscript, did the cluster analysis, and provided figures. M. Pleyer contributed to conceiving and designing the analysis, collecting and coding the data, and conceptualizing and writing the manuscript. M.S. contributed to the coding and to conceptualizing and writing the manuscript. M. Placiński contributed to the coding and to conceptualizing and writing the manuscript, did the cluster analysis, and prepared the figures. P.Z. contributed to the coding and to conceptualizing and writing the manuscript. S.W. contributed to conceiving and designing the analysis, collecting and coding the data, and conceptualizing and writing the manuscript. All authors reviewed the manuscript.

Funding This research was supported by the Polish National Science Centre under grant agreement UMO-2019/34/E/HS2/00248. M. Pleyer was supported under grant agreement UMO-2021/43/P/HS2/02729 co-funded by the National Science Centre and the European Union Framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie grant agreement No. 945339. For the purpose of Open Access, the authors have applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

Declaration

Conflict of interest The authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Bar-On, D. (2021). How to do things with nonwords: Pragmatics, biosemantics, and origins of language in animal communication. *Biology & Philosophy*, *36*(6), 1–25.

Bartlett, F. C. (1932). *Remembering. A study in experimental and social psychology*. Cam- bridge University Press.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acous- tical Society of America*, *22*(6), 725–730.

Bavelas, A. (1952). Communication patterns in problem-solving groups. In H. von Foerster (Ed.), *Cybernetics. Circular causal and feedback mechanisms in biological and social sys- tems. Transactions of the ninth conference* (pp. 1–44). Josiah Macy Jr. Foundation New York, NY.

Blokpoel, M., van Kesteren, M., Stolk, A., Haselager, P., Toni, I., & Van Rooij, I. (2012). Recipient design in human communication: Simple heuristics or perspective taking? *Frontiers in human neuroscience*, *6*, 253.

Caldwell, C. A., & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *Plos One*, *7*(8), e43807. https://doi.org/10.1371/journal.pone.0043807.

Campbell, R., & Grieve, R. (1981). Royal Investigations of the Origin of Language. *Historiographia Linguistica*, *9*, 43–74. https://doi.org/10.1075/hl.9.1-2.04cam.

Christiansen, M. H., & Chater, N. (2022). *The Language game: How Improvisation created Language and changed the World*. Penguin Press.

Cuskley, C. (2019). Alien forms for alien language: Investigating novel form spaces in cultural evolution. *Palgrave Communications*, *5*(1), 1–15.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation. *Behavioural ecology: An evolutionary approach*, *2*, 282–309.

dos Santos, M., Rodrigues, J. F. M., Wedekind, C., & Rankin, D. J. (2012). The establishment of communication systems depends on the scale of competition. *Evolution and Human Behavior*, *33*(3), 232–240.

Esper, E. A. (1925). *A technique for the experimental investigation of associative interference in artificial linguistic material*. Linguistic Society of America.

Fay, N., Walker, B., & Swoboda, N. (2017). Deconstructing Social Interaction: The Complimentary Roles of Behaviour Alignment and Partner Feedback to the Creation of Shared Symbols. In *CogSci*.

Ferretti, F. (2022). *Narrative persuasion. A cognitive perspective on Language Evolution* (7 vol.). Springer Nature.

Ferretti, F., & Adornetti, I. (2021). Persuasive conversation as a new form of communication in Homo sapiens. *Philosophical Transactions of the Royal Society B*, *376*(1824), 20200196.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, *29*(5), 737–767.

Galantucci, B. (2017). Experimental semiotics. In *Oxford Research Encyclopedia of Linguistics*. Retrieved 2 March 2023 from https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-210. https://doi.org/10.1093/acrefore/9780199384655.013.210

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: A review. *Frontiers in human neuroscience*, *5*, 11.

Galantucci, B., & Roberts, G. (2012). Experimental Semiotics: An engine of discovery for understanding human communication.Advances in Complex Systems, *15* (03n04).

Galantucci, B., Garrod, S., & Roberts, G. (2012a). Experimental semiotics. Language and Linguistics Compass, 6(8), 477–493.

Galantucci, B., Theisen, C., Gutierrez, E. D., Kroos, C., & Rhodes, T. (2012b). The diffusion of novel signs beyond the dyad. Language Sciences, 34(5), 583–590.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, *31*(6), 961–987.

Iizuka, H., Marocco, D., Ando, H., & Maeda, T. (2013). Experimental study on co-evolution of categorical perception and communication systems in humans. *Psychological research*, *77*(1), 53–63.

Inoue, N., & Morita, J. (2021). A behavioral task for exploring dynamics of communication system in dilemma situations. *Artificial Life and Robotics*, *26*, 329–337.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of personality and social psychology*, *4*(3), 343.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and Manipulation. *Behavioral Ecology: An Evolutionary Approach* (pp. 380–402). Blackwell.

Lee, J. J., & Pinker, S. (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review*, *117*(3), 785.

Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

Morris, C. W. (1938). Foundations of the theory of Signs. *International encyclopedia of unified science* (pp. 1–59). Chicago University Press.

Motamedi, Y., Smith, K., Schouwstra, M., Culbertson, J., & Kirby, S. (2018). The emergence of systematic argument distinctions in artificial sign languages.

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition*, *192*, 103964.

Nielsen, F., & Nielsen, F. (2016). Hierarchical clustering. *Introduction to HPC with MPI for Data Science*, 195–211.

Nölle, J., & Galantucci, B. (2021). Experimental semiotics: Past, present, and future. In A. M. García, & A. Ibâñez (Eds.), *The Routledge Handbook of Semiosis and the brain* (pp. 66–81). Routledge.

Nölle, J., Staib, M., Fusaroli, R., & Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, *181*, 93–104.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Perlman, M., & Lupyan, G. (2018). People can create iconic vocalizations to communicate various meanings to naïve listeners. *Scientific reports*, *8*(1), 2634.

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society open science*, *2*(8), 150152.

RapidTables. (n.d.). https://www.rapidtables.com/

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Compositional structure can emerge without generational transmission. *Cognition*, *182*, 151–164.

Roberts, G., & Clark, R. (2020). Dispersion, communication, and alignment: An experimental study of the emergence of structure in combinatorial phonology. *Journal of Language Evolution*, *5*(2), 121–139.

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

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences*, *104*(18), 7361–7366.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359–393.

Stevens, J. S., & Roberts, G. (2019). Noise, economy, and the emergence of information structure in a laboratory language.Cognitive Science, 43(2), e12717.

Tamariz, M. (2017). Experimental studies on the cultural evolution of language. *Annual Review of Linguistics*, *3*, 389–407.

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.

Van Rossum, G., & Drake, F. L. Jr. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Wolfle, D. L. (1933). The relative Stability of First and Second Syllables in an Artificial Language. *Language*, *9*(4), 313–315. https://doi.org/10.2307/409418.

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of planning education and research*, *39*(1), 93–112.

Zlatev, J., Wacewicz, S., Zywiczynski, P., & van de Weijer, J. (2017). Multimodal-first or pantomime-first?: Communicating events through pantomime with and without vocalization. *Interaction Studies*, *18*(3), 465–488.

Zubek, J., Korbak, T., & Rączaszek-Leonardi, J. (2023). Models of symbol emergence in communication: a conceptual review and a guide for avoiding local minima. arXiv:2303.04544. https://doi.org/10.48550/arXiv.2303.04544

Żywiczyński, P. (2018). *Language Origins: From mythology to science*. Peter Lang.

Żywiczyński, P., Sibierska, M., Wacewicz, S., van de Weijer, J., Ferretti, F., Adornetti, I., & Deriu, V. (2021). Evolution of conventional communication. A cross-cultural study of pantomimic re-enactments of transitive events. *Language & Communication*, *80*, 191–203.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Angelo Delliponti[1] · Renato Raia[2] · Giulia Sanguedolce[2] · Adam Gutowski[1] · Michael Pleyer[1] · Marta Sibierska[1] · Marek Placiński[1] · Przemysław Żywiczyński[1] · Sławomir Wacewicz[1]

✉ Angelo Delliponti
503302@doktorant.umk.pl

Renato Raia
raiarenato@gmail.com

Giulia Sanguedolce
sangue.giulia@gmail.com

Adam Gutowski
adgut1509@gmail.com

Michael Pleyer
pleyer@umk.pl

Marta Sibierska
sibier@umk.pl

Marek Placiński
marpla@umk.pl

Przemysław Żywiczyński
Przemyslaw.Zywiczynski@umk.pl

Sławomir Wacewicz
wacewicz@umk.pl

1 Nicolaus Copernicus University in Toruń, Toruń, Poland

2 Roma Tre University, Rome, Italy

3. Appendix

3.1. Which mindreading for ostensive communication? An event-related potentials study of how the brain processes communicative and informative intentions

Cognitive Science

Which mindreading for ostensive communication? An event-related potentials study of how the brain processes communicative and informative intentions

--Manuscript Draft--

Manuscript Number:	
Full Title:	Which mindreading for ostensive communication? An event-related potentials study of how the brain processes communicative and informative intentions
Article Type:	Regular Article
Keywords:	cognitive pragmatics; event-related potentials; intentions; mindreading; N170; ostensive communication; Relevance Theory
Corresponding Author:	Francesco Ferretti Roma Tre University: Universita degli Studi Roma Tre Rome, ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Roma Tre University: Universita degli Studi Roma Tre
Corresponding Author's Secondary Institution:	
First Author:	Francesco Ferretti
First Author Secondary Information:	
Order of Authors:	Francesco Ferretti
	Angelo Delliponti
	Valentina Deriu
	Alessandra Chiera
	Daniela Altavilla
	Serena Nicchiarelli
	Sławomir Wacewicz
	Ines Adornetti
Order of Authors Secondary Information:	
Abstract:	According to the ostensive-inferential model, human communication is characterized by two different types of intentions: communicative intention (CI) and informative intention (II). In its classical formulation, the dual level of intentions that characterizes ostensive communication (OC) is considered an exclusive prerogative of adult humans. In recent years, a deflationist perspective on OC has emerged, challenging the prevailing classical view that OC is an all-or-nothing phenomenon. This new approach suggests that basic forms of OC can be observed in both human infants and non-human primates. A pivotal aspect of the ongoing debate focuses on the nature of the mindreading process underlying OC. Classical perspectives posit the hypothesis of high-level inferential mindreading for both CI and II processing. Conversely, deflationary perspectives associate basic forms of mindreading with basic forms of OC. To adjudicate between these two models, we present an event-related potentials (ERPs) study on the time course of processing communicative and informative intentions expressed through a combination of eye contact and gestures. Three primary findings emerged from the analysis of the ERPs, relating to the amplitude of two early components, i.e., P100 and N170, and one later component, i.e., LC1 (600-800 ms). Overall, the findings suggest that the detection of both communicative and informative intentions occurs within the 200-millisecond window, favoring conceptions of mindreading grounded in low-level rather than high-level cognitive processes. We address the empirical and theoretical implications of these findings within the context of a deflationary perspective on OC.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Roma Tre

Rome, 22 January 2025

Dear Editor,

We are pleased to submit our article “Which mindreading for ostensive communication? An event-related potentials study of how the brain processes communicative and informative intentions” for being considered for its publication in *Cognitive Science*.

The article reports on an electroencephalography study of the neural correlates of the processing of ostensive communication. We believe that *Cognitive Science* is the most appropriate journal because our research has important implications for multiple disciplines, including:

- Cognitive pragmatics: our results contribute to the Relevance Theory debate.
- Neuroscience: our data shed light on the neural timeline of mindreading.
- Cognitive Psychology: our data shed light on the nature and kind of mindreading involved in ostensive communication.
- Developmental psychology: our data are consistent with experimental research showing that basic forms of ostensive communication emerge early in ontogeny.
- Comparative psychology: our data are consistent with experimental research showing that basic forms of ostensive communication are also present in non-human primates.
- Language evolution research: our data support a pantomimic scenario for the origin of language.

On a different note, we wish to state that:

- our manuscript meets the guidelines for ethical conduct and report of research;
- no potential or actual conflicts of interest exist regarding our research;
- our manuscript and data have not been published previously and that they are not under consideration for publication elsewhere;
- the datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request;
- all authors have contributed significantly to the manuscript and consent to their names on the paper.

Thank you very much for considering our submission.

Best regards,

Francesco Ferretti & Co-authors

Which mindreading for ostensive communication? An event-related potentials study of how the brain processes communicative and informative intentions

Francesco Ferretti[1,]*, Angelo D. Delliponti[2], Valentina Deriu[1], Alessandra Chiera[1], Daniela Altavilla[1], Serena Nicchiarelli[1], Sławomir Wacewicz[2], and Ines Adornetti[1]

[1]Cosmic Lab, Department of Philosophy, Communication and Performing Arts, Roma Tre University, Rome, Italy.

[2] Center for Language Evolution Studies, Nicolaus Copernicus University in Toruń, Toruń, Poland

IMPORTANT NOTE: Francesco Ferretti and Angelo D. Delliponti share the first authorship of this study.

***Corresponding author:**
Prof. Francesco Ferretti
Cosmic Lab, Department of Philosophy, Communication and Performing Arts
Roma Tre University,
Via Ostiense 236
00134, Rome, Italy
francesco.ferretti@uniroma3.it
ORICID: https://orcid.org/0000-0003-2109-7297

Co-authors email addresses:

Angelo D. Delliponti: angelo.d.delliponti@gmail.com

Valentina Deriu: valentina.deriu@uniroma3.it

Alessandra Chiera: alessandra.chiera@uniroma3.it

Daniela Altavilla: daniela.altavilla@uniroma3.it

Serena Nicchiarelli: serena.nicchiarelli@uniroma3.it

Sławomir Wacewicz: wacewicz@umk.pl

Ines Adornetti: ines.adornetti@uniroma3.it

Declarations

Data availability statements

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Conflict of interest

No potential conflict of interest was reported by the author(s).

Ethics approval statement

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The current study was approved by the Ethics Committee of Roma Tre University, Rome.

Author Contributions

FF planned the study, supervised the recruitment of participants and administration of the task, interpreted the results, and wrote the paper. ADD planned the study, contributed to the preparation of the experimental materials, recruited the participants, administered the task, contributed to the interpretation of the results, contributed to the final draft. VD contributed to the design of the study, to the preparation of the experimental materials, processed the data, contributed to the interpretation of the results and to the final draft. AC contributed to the design of the study, to the preparation of the experimental materials, to the interpretation of the results and to the final draft. DA contributed to the design of the study, to the processing of the data, to the interpretation of the results and to the final draft. SN drew the figures and contributed to the interpretation of the results. SW contributed to the design of the study, to the interpretation of the results, and to the final draft. IA contributed to the design of the study, supervised the recruitment of participants and administration of the task, contributed to the interpretation of the results and to the final draft.

Anonymous references

Anonymous 1 et al., 2022: Altavilla, D., Adornetti, I., Chiera, A., Deriu, V., Acciai, A., & Ferretti, F. (2022). Introspective self-narrative modulates the neuronal response during the emphatic process: An event-related potentials (ERPs) study. *Experimental Brain Research*, *240*(10), 2725-2738. https://doi.org/10.1007/s00221-022-06441-4

Anonymous 2 et al., 2022: Chiera, A., Adornetti, I., Altavilla, D., Acciai, A., Cosentino, E., Deriu, V., ... & Ferretti, F. (2022). Does the character-based dimension of stories impact narrative processing? An event-related potentials (ERPs) study. *Cognitive Processing*, *23*(2), 255-267. https://doi.org/10.1007/s10339-021-01070-1

Anonymous 3 et al., 2013: Cecchini, M., Aceto, P., Altavilla, D., Palumbo, L., & Lai, C. (2013). The role of the eyes in processing an intact face and its scrambled image: A dense array ERP and low-resolution electromagnetic tomography (sLORETA) study. *Social Neuroscience*, *8*(4), 314–325. https://doi.org/10.1080/17470919.2013.797020

Anonymous 4 et al., 2023: Adornetti, I., Chiera, A., Deriu, V., Altavilla, D., & Ferretti, F. (2023). Comprehending stories in pantomime. A pilot study with typically developing children and its implications for the narrative origin of language. *Language & Communication*, *93*, 155-171. https://doi.org/10.1016/j.langcom.2023.10.001

Anonymous 5 2022: Ferretti, F. (2022). *Narrative persuasion. A cognitive perspective on language evolution*, Springer Nature, Cham.

Anonymous 6 et al., 2020: Zlatev, J., Żywiczyński, P., & Wacewicz, S. (2020). Pantomime as the original human-specific communicative system. *Journal of Language Evolution*, *5*(2), 156-174. https://doi.org/10.1093/jole/lzaa006

Manuscript

Which mindreading for ostensive communication? An event-related potentials study of how the brain processes communicative and informative intentions

Abstract

According to the ostensive-inferential model, human communication is characterized by two different types of intentions: communicative intention (CI) and informative intention (II). In its classical formulation, the dual level of intentions that characterizes ostensive communication (OC) is considered an exclusive prerogative of adult humans. In recent years, a deflationist perspective on OC has emerged, challenging the prevailing classical view that OC is an all-or-nothing phenomenon. This new approach suggests that basic forms of OC can be observed in both human infants and non-human primates. A pivotal aspect of the ongoing debate focuses on the nature of the mindreading process underlying OC. Classical perspectives posit the hypothesis of high-level inferential mindreading for both CI and II processing. Conversely, deflationary perspectives associate basic forms of mindreading with basic forms of OC. To adjudicate between these two models, we present an event-related potentials (ERPs) study on the time course of processing communicative and informative intentions expressed through a combination of eye contact and gestures. Three primary findings emerged from the analysis of the ERPs, relating to the amplitude of two early components, i.e., P100 and N170, and one later component, i.e., LC1 (600-800 ms). Overall, the findings suggest that the detection of both communicative and informative intentions occurs within the 200-millisecond window, favoring conceptions of mindreading grounded in low-level rather than high-level cognitive processes. We address the empirical and theoretical implications of these findings within the context of a deflationary perspective on OC.

Keywords: cognitive pragmatics; event-related potentials; intentions; mindreading; N170; ostensive communication; Relevance Theory.

1. Introduction

In models of cognitive pragmatics inspired by Grice (1957, 1989), human communication is essentially the expression and understanding of intentions. In this sense, Sperber and Wilson (1986/1995) argue that "communication is successful not when hearers recognize the linguistic meaning of the utterance, but when they infer the speaker's "meaning" for it" (Sperber and Wilson, 1986/1995, p. 23). In such a perspective, what the speaker intends to say (the informational content he/she intends to express) through an utterance is not the only intention that the listener needs to grasp in order to understand the meaning of that utterance: it is also necessary for the listener to understand that the speaker intends to communicate that informative content to him/her. The two authors call this model *ostensive-inferential* communication, which is characterized by two different types of intentions: informative intention and communicative intention. As highlighted by Scott-Phillips (2015a), the "ostensive" nature of ostensive-inferential communication is related to the fact that "meaningful communication is not only intentional, it is also *overtly* intentional – it brings attention to the intentions that are being expressed" (Scott-Phillips, 2015b, pp. 802-803).

The dual level of intentions that characterizes ostensive communication (OC, thereafter) has made this form of communication (in its classical formulation) an exclusive prerogative of humans (e.g., Origgi and Sperber, 2000; Sperber and Origgi, 2010). One of the ways in which proponents of the classical thesis justify the qualitative difference between human and animal communication is by identifying cognitive architectures exclusive to humans capable of processing the dual level of intentions involved in ostensive communication. In a communicative system based on the expression and understanding of intentions, the mindreading system is the main player: both informative and communicative intentions require a system allowing the representation of the minds of others. The link between mindreading and OC is so strong that it has led some authors to argue for the existence of mindreading as a necessary condition for the emergence and functioning of human communication (Sperber, 2000; Origgi and Sperber, 2000; Tomasello, 2008). Indeed, as Scott-Phillips (2015a, p. 68) argues, "without it, there could be no ostensive communication, and hence no linguistic communication, and no languages".

In light of these considerations, the study of the cognitive architectures involved in the processing of informative and communicative intentions becomes an essential step in investigating the nature of OC. As will be argued, this is an essential step in answering the question of whether this form of communication is a trait specific to our species (according to an all-or-nothing logic), or whether simpler forms of OC can also be attributed to nonhuman animals, according to a gradualist perspective. In line with these considerations, the basis of the present study is the investigation of the type (or types) of mindreading that can be assumed to underlie the functioning of OC.

1.1 The Classical Perspective

The first question to be analyzed in this context is whether it is sufficient to refer to a generic type of mindreading that is generally used to interpret the actions of others, or whether OC instead requires specialized devices for linguistic communication that are designed to recognize and understand the speaker's intentions. While it is reasonable to speculate that the early stages of our ancestors' OC was founded on a generic mindreading ability that enabled them to interpret the actions of others (Sperber, 2000; Origgi and Sperber, 2000), the prevailing view is that reference to the speaker's intentions (both communicative and informative intentions) brings specialized processing devices into play (Sperber and Wilson, 2002; Scott-Phillips, 2015a). Sperber and Wilson (2002) argue that speaker intentions are different from general intentions to behave in a certain way, and that in order to cope with this difference, the devices underlying the reading of speaker intentions must have special features that distinguish them from generic mindreading systems. More specifically, since ostensive-inferential communication only emerges as a result of the processing of numerous levels of mindreading, Scott-Phillips (2015a) claims that recursive mindreading systems must be brought into play. Indeed, since OC "involves not just the mental representation of others' mental states, but the mental representation of others' mental representations of one's own mental states, and indeed several further levels of representation beyond this" (Scott-Phillips, 2015a, p. 64-65), what is needed is a system capable of processing multiple recursively embedded levels of mindreading, i.e., a high-level processing system.

In arguments that support the existence of a qualitative difference between human and animal communication, reference to high-level cognitive architectures has been identified as a crucial component. Indeed, within the classical hypothesis (Sperber and Wilson, 1986/1995; 2002;

Scott-Phillips, 2015a; 2015b), the notion of high models of mindreading lends support to the view that OC constitutes the specific trait of human communication. Absent the capacity for recursive mindreading, nonhuman animals are constrained to a mode of communication that can be characterized as the code model (Shannon and Weaver, 1948). In this view, the difference between human and animal communication rests on the fact that OC "depends upon sophisticated forms of social cognition that are unique to humans, and which evolved in our species as a result of our über-social nature" (Scott-Phillip, 2015a, p. xiii).

Such a position underlies an all-or-nothing conception of OC: there are no different degrees of OC capable of bridging the distance between animal and human communication. From this point of view, the ostensive model of communication (in its classical version) appears to be plagued by a form of discontinuity, a major difficulty in a perspective set within an evolutionary framework. In contrast to the discontinuity thesis, Moore (in press) proposes a gradualist approach based on the possibility that even nonhuman animals can be ascribed forms of OC (albeit to varying degrees).

1.2. A deflationary proposal

In light of the considerations that have been raised, the issue of the role of mindreading in OC can be further refined into a more specific inquiry: is it possible to hypothesize a more basic level of ostensive communication based on mindreading mechanisms that are less elaborate than the recursive mindreading (defended by proponents of the classical thesis)? The point here is twofold: to characterize minimal forms of mindreading; to show that these minimal forms of mindreading are sufficient to guarantee basic ostensive forms of communication. In other words, the task at hand is to lower the bar for both the properties of OC and the systems assigned to its processing.

Moore (in press) cites several authors who argue for a less cognitively demanding OC (Gómez, 1994; Brink, 2004; Moore, 2016, 2017). In particular, and this is the nodal point, his idea is that it is possible to satisfy Grice's clause regarding communicative intentions (Clause II) without having to invoke meta-representational processing systems. Moore's view is that it is sufficient to refer to processes of mutual attention to capture communicative intention: "if speakers deliberately address their utterances to an intended audience, as a way of getting intended recipient to respond to the utterance, then such speakers would satisfy Grice's second clause intention" (Moore, in press, p. 29). More specifically:

> The act of address serves to make the speaker utterance open (or “ostensive”) in the way that suffice to qualify it as communicative: moreover, knowing how to address utterances requires neither metarepresentational abilities, nor the ability to think about others’ mental states. Instead, (…) the process requires only mastery of embodied forms of address that constitutes a basic feature of interactions between agents (Moore, in press, p. 29).

Considering communicative intention in terms of an act of address makes it possible to consider OC as a matter of degree (rather than quality), and thus makes it plausible to assume ostensive forms of communication even in nonhuman animals. Deflationary perspectives nowadays represent a point of convergence of cognitively oriented post-Gricean studies (for a proto-Gricean alternative to Moore's minimally Gricean proposal, see also Bar-On, 2021). The common intent is to search for basic forms of OC that potentially represent both an early stage of development of human communication in ontogeny and a link to some forms of non-human communication. Even some of the authors who have most tenaciously defended the thesis of a qualitative difference between ostensive and animal communication in the past are now proposing a revision of their theories. Scott-Phillips and Heintz (2023), for example, argue for a gradualist model of meaning based on the distinction of three different forms of social relations (specifically, attention manipulation), insisting that "these three varieties of meaning are special cases of one another, and as such they are a key source of evolutionary gradualism between humans and other great apes" (Scott-Phillips and Heintz, 2023, p. 8).

For the purposes of the present study, it is noteworthy that a significant component of Scott-Phillips and Heintz's (2023) revision of the model of communication (and meaning) is closely related to the revision of the cognitive processes and systems involved in ostensive communication. With reference to processes of "attention manipulation", they distinguish between Ladyginian and Gricean forms of communication. In the former, "individuals intentionally manipulate others’ attention toward evidence of their (the focal individual’s) own intentions" (Scott-Phillips and Heintz, 2023, p. 2); in the latter, "individuals intentionally manipulate others’ attention toward evidence of a specific type of intention, namely informative intentions" (ibid.). In the present investigation, the forms of communication under consideration are of a Ladyginian nature, as they represent a potential bridge to certain forms of animal communication.

In revising some core tenets of Relevance Theory, Sperber and Wilson (2024) also argue that studies of nonhuman animals and those related to child developmental psychology support the existence of basic forms of OC, thus challenging the all-or-nothing thesis that characterized their classic proposal from the first publication of Relevance Theory (Sperber and Wilson, 1986/1995) until recently (Sperber and Wilson, 2002; Sperber and Origgi, 2010). The idea that OC should be considered in a gradualist framework is achieved by recognizing a basic (non-mentalist) form of OC. Here is the summary framework of the model proposed by the two authors (Sperber and Wilson, 2024):

> INFORMATIVE INTENTION: The intention to make some specific information manifest to another.
>
> COMMUNICATIVE INTENTION (in ostensive communication in general): The intention to make manifest to the addressee that the communicator is addressing him with an informative intention.
>
> COMMUNICATIVE INTENTION (in mentalistic ostensive communication): The intention to make manifest to the addressee that the communicator is addressing him with the intention of informing him THAT________.

Once more, the examination of the nature of OC is inextricably linked to the study of underlying cognitive processes and systems. Specifically, metarepresentational and inferential processes are observed. In a perspective in line with Moore's deflationary proposals, Sperber and Wilson argue that the answer to the question "Does the addressee of an act of ostensive communication have to draw a mentalistic inference to understand that the communicator is acting with the intention of addressing him?" must be "Not necessarily" (Sperber and Wilson, 2024, pp. 7-8). In other words, it is not necessary to invoke complex cognitive architectures, such as recursive mindreading, to account for basic forms of communicative intention. Indeed, *eye contact* constitutes the primary modality through which individuals address one another.

> In non-verbal communication, whether among human or other apes, there are several ways of indicating to the addressee that he is being addressed: in particular by establishing eye contact, which is prototypical signal of a communicative intent (Csibra and Gerely, 2009; Gòmez, 1996; Moore, 2016). So ostension is a type of action that can be recognized without metarepresenting the communicator's

> intention, as the study of ostension addressed to infants has shown (Csibra, 2010). (Sperber and Wilson, 2024, p. 9).

As discussed in the following sections, the present experimental study places particular emphasis on eye contact. In this section, it is sufficient to state that the reference to less sophisticated processing systems than those assumed in the classical version of relevance theory underlies a review of the peculiarities of OC. In particular, the all-or-nothing conception will be challenged in favor of a gradualist perspective for the emergence of OC both in ontogeny and phylogeny. As for ontogeny, young infants have been shown to be sensitive to various ostensive cues such as eye gaze (Senju and Csibra 2008; Senju et al., 2008; Loria 2017). Experimental studies indeed reveal that infants as young as 12 months are able to understand ostensive communicative acts (e.g., Liszkowski et al., 2008; Schulze and Tomasello, 2015; Siposova et al., 2018). For example, Schulze and Tomasello (2015) show that 18-month-old children are able to understand indirect ostensive communicative acts in which the adult's communicative intention is conveyed through gaze. In terms of phylogeny, as Moore (2016) notes, great apes, particularly chimpanzees, employ eye contact during gestural communication as a means to solicit the attention of their audience.

As previously discussed, the reference to act of address or attention manipulation is a critical component in the arguments of theoretical models that argue for fundamental forms of OC. How can such processes be conceptualized within the broader framework of the relationship between cognitive architectures and OC? Assuming that such processes are not inferential and meta-representational, does this imply that mindreading is entirely excluded from basic forms of OC? In other words, the central question is whether basic ostensive communication precludes mindreading entirely, or whether it is feasible to hypothesize basic forms of mindreading capable of elucidating low-level forms of OC. Given the close interdependence between mindreading and OC, determining whether mindreading is also implicated in basic forms becomes a pivotal issue for this discussion.

1.3. Which mindreading for a deflationary model of OC?

In the debate about the nature of mindreading, it is common to refer to two families of models. Theory-Theory (TT) includes those (high-level) models of mindreading that consider the interpretation of behavior in terms of the psychology of commonsense beliefs and desires, the idea that people understand the actions of others by attributing propositional attitudes to other agents using a true "theory" of mind (an organized belief system) (e.g., Fodor, 1987; Davies and Stone, 1995a). An alternative way to explain mindreading is the so-called simulation theory (ST) (e.g., Davies and Stone, 1995b; Gallese and Sinigaglia, 2011; Goldman, 2006): proponents of this alternative account argue that there is no need to refer to a theory to explain the behavior of others "because we have a model, our own mind, that we can use to simulate the other person's mental states" (Gallagher, 2008, p. 535).

Although the discovery of mirror neurons provided a crucial impetus for ST (e.g., Gallese and Goldmann, 1998; Gallese 2013), a view is that the two models are not mutually exclusive but represent different ways of interpreting behavior that serve different functional needs depending on the task at hand (e.g., Bohl and van den Bos, 2012; Keysers and Gazzola, 2007). However, a salient point for the purposes of the arguments discussed herein is the notion that, in contrast to TT (which is a third-person perspective), ST is a first-person perspective capable of dispensing with metarepresentational structures and inferential processing. And it is largely because of this aspect that ST has traditionally been considered a more basic mindreading system than TT.

Can the mindreading devices involved in basic forms of OC be thought of in simulationist terms? To our knowledge, the only study (theoretical and not empirical) that attempts to answer this question is that of Yousefi Heris (2020). However, his conclusion is negative: simulationist models alone cannot succeed in the task of identifying the speaker's communicative intentions; only a mixed perspective (ST and TT) can do so. Nevertheless, Yousefi Heris' response does not take into account the question of basic forms of ostensive communication. Instead, it considers a full-fledged thesis of pragmatic competence. Therefore, it is not possible to determine whether simulationist models can be used for basic models of OC, as Haris's arguments do not address this issue. In the present research, we therefore aim to fill this gap by attempting to understand whether basic forms of mindreading underlie basic forms of OC.

1.4 Existing EEG studies on the time course of mindreading

In their revision of Relevance Theory, Sperber and Wilson (2024) posit two levels of ostensive communication: the first non-mentalistic, the second mentalistic. According to this hypothesis, the initial level, corresponding to the communicative intention of the first type, does not necessitate forms of mindreading but rather facilitates the subsequent activation of mentalization, which is essential for comprehending the informational content inherent to the communicative intention of the second type. From this point of view, informative intention is crucial for the activation of mindreading processes and for characterizing communication as fully ostensive. This two-level hypothesis appears to be consistent with the notion of a two-stage process of mindreading. In fact, a suitable approach for examining this issue empirically is to assess the time course of mindreading processes involved in understanding ostensive communicative acts. One experimental methodology that has the capacity to accomplish this is the electroencephalography (EEG), specifically the analysis of event-related potentials. Accordingly, we present an explorative EEG study to assess the nature of mindreading involved in ostensive communication, by examining the timeline of brain processes involved in understanding communicative and informative intentions.

To the best of our knowledge, there are no specific EEG studies on the role of the mindreading timeline in ostensive communication comprehension tasks. Among the limited extant works related to the timeline of mindreading in general (e.g. Sabbagh et al., 2004; Ruzzante and Vaes 2021), some EEG studies appear to support the two-stage processing thesis. Ruzzante and Vaes' (2021) examined the mindreading timeline in a facial expression recognition task. The results of their EEG experiment support the hypothesis of two-stage processing, comprising a preliminary perceptual stage and a subsequent contextual stage. Specifically, Ruzzante and Vaes (2021) identify an initial stage, corresponding to the N170 component, of perceptual configuration that marks the onset of a *mind detection* phase, in which "people detect the presence of a mind for the first time marking the beginning of the mentalization process" (ibid., p. 11), and a later second stage, corresponding to the P300 component, of contextual information integration that characterizes a *mind attribution* phase. What happens in the second stage "was influenced by both perceptual and contextual information, allowing us to conclude that P300 marks the second phase in the timeline of the mentalization process, where both perceptual and contextual information are integrated to attribute a mind to others" (ibid., p. 11). The experimental results obtained by Ruzzante and Vaes (2021) do not concern the ostensive nature of communication. However, they seem to be in line with the two-stage thesis of the models of OC that adopt a deflationist perspective. In particular, it is possible to interpret the stage of mind detection as related to the level of communicative intention processing, while the stage of mind attribution, which is concerned with content attribution, is related to the level of informative intention processing.

1.5.The present study

To ascertain whether the two-stage model genuinely constitutes the most congruent approach with a deflationary perspective on OC, in the present study we have constructed an experimental task that directly aimed to investigate the time course of processes involved in understanding basic forms of OC. Specifically, event-related potentials (ERPs) are recorded in response to visual stimuli in which communicative and informative intentions were manipulated. The stimuli are constructed to generate expectations about the successful outcome of the ostensive communicative act. Participants observe visual stories in which one of the depicted characters makes a request (e.g., pass an object, be quiet, dance together) to a second agent. In making this request, the first character manifests a communicative intention and an informative intention, which were expressed through a combination of eye contact and gestures, respectively. Three conditions are compared: congruent, in which both intentions are satisfied; semi-congruent, in which the informative intention, but not the communicative intention, is satisfied; and incongruent, in which neither the communicative intention nor the informative intention are satisfied. This manipulation allows us to observe how the two types of intentions are processed.

To support the two-stage thesis, we should observe that the processing of communicative and informative intentions occurs at different times (an earlier and a later one). Assuming that communicative intention (Sperber and Wilson's first level) corresponds to the stage of mind detection, we analyze the earlier ERP component, i.e. the N170; while for the informative intention (Sperber and Wilson's second level), which should correspond to the second stage of mind attribution, we analyze the later ERP components, i.e., the P300 and late components.

As for the N170, we hypothesize that we will find a greater amplitude in the response to the semi-congruent and incongruent conditions compared to the congruent condition, indicating that the violation of communicative intention has been detected. As for the amplitude of the P300, we hypothesize that we will not find significant differences between the congruent and semi-congruent conditions, since the informative intention is satisfied in both conditions; on the contrary, we should observe a greater P300 amplitude in response to the incongruent condition compared to the congruent and semi-congruent conditions, suggesting a contextual information update due to the violation of the informative intention.

2. Materials and methods

2.1 *Participants*

Forty-nine participants were recruited for this study. All participants were right-handed, native Italian speakers, had no neurological and/or psychiatric diagnoses, were not taking any medications and/or drugs, and had normal or corrected-to-normal vision. After cleaning the electroencephalographic data, 43 participants were included in the statistical analysis (22 females; mean age 27,09; SD 6,52). Six participants were excluded due to the presence of too many artifacts in the EEG registration. The present study was approved by the Ethical Committee of [masked information for blind review].

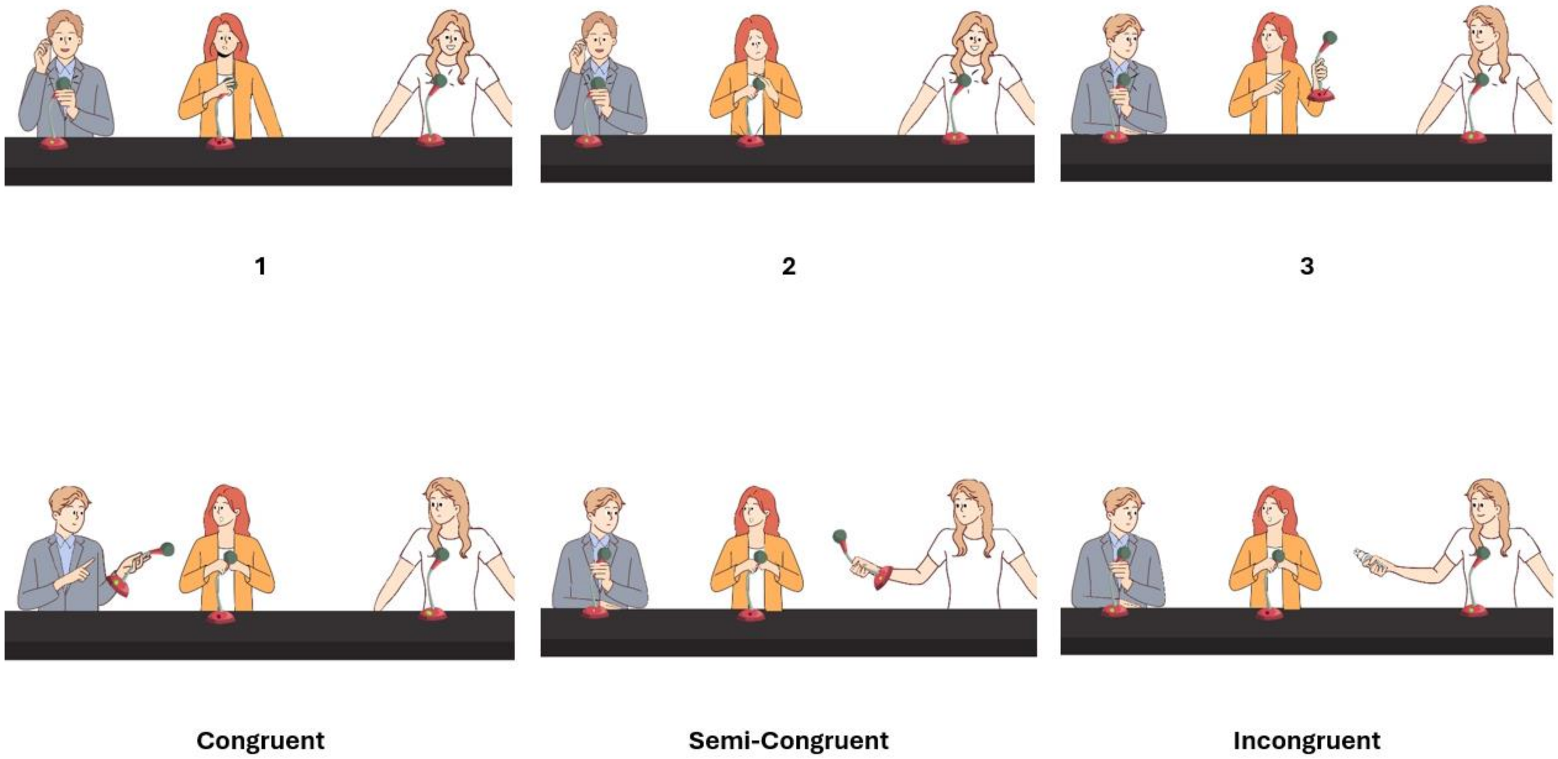
2.2 Stimuli

Stimuli consisted of 90 (30 per condition) visual stories presented in three experimental conditions: Congruent (CONG); Semi-Congruent (SEMI-CONG); Incongruent (INCONG). The stories consisted of four colored pictures representing OC between two characters, a communicator X and a receiver Y, in which X intentionally communicates information to Y through a combination of eye contact and gestures. Another character, Z, is present in the scene as an observer who is not directly involved in the communicative interactions. In each stimulus, the first scene introduced the OC general neutral context and the three characters (the main character X was always positioned in the middle between the other two). The second picture depicted an event that occurred to character X, followed by a scene (third picture) in which X intentionally communicates information (a specific request, e.g., for help) to character Y through a combination of eye contact, which conveys the communicative intention (CI), and a gesture (e.g., pointing), which conveys the informative intention (II). It is important to emphasize that in this third scene, both the receiver Y and the observer Z are looking at the request of X, but the gaze of the communicator X is directed at the receiver Y in order to communicate with her/him. This was also represented in the picture by the head of X turning in the direction of Y. Finally, the last (i.e. fourth) picture changed according to the experimental condition: in the congruent condition, the scene depicted an interaction in which the receiver Y responds to the request of X in such a way that both the CI and the II are satisfied; in the semi-congruent condition, the II of the communicator X is satisfied, but not his/her CI, i.e., Z satisfies X's request; in the incongruent condition, neither the CI nor the II are satisfied, i.e., the character Z responds to X's communication in an incongruent way. An example of an experimental stimulus is shown in Figure 1.

The 90 stories were organized into three counterbalanced lists (List A, List B, and List C) to ensure that each story appeared in only one condition per list and per participant. Thus, List A contained 90 visual stories (30 CONG; 30 SEMI-CONG; 30 INCONG), whereas Lists B and C were obtained by replacing each condition with one of the remaining two conditions. In addition, each of the three lists also contained 15 fillers, i.e., a color picture of an object (e.g., tennis racket), so that participants received a total of 105 stimuli. For each participant, stories were presented from either List A, List B, or List C.

The recording of ERPs was time-locked to the onset of the target picture, which was always the last (fourth) picture.

Figure 1. In the first scene (1), the image shows three agents in the context of a conference, sitting at a table with a microphone in front of each of them. In the second image (2), the woman in the center (X) starts to speak, but taps the microphone because it is not working. In the third image (3), the communicator (X) raises the microphone pointing to the object (II) and turns her gaze to the man on her left, the receiver (Y), asking for help in solving the problem with the microphone. In this scene, the character to the left of the communicator (Z) also observes the request, but the eye contact is only between the communicator (X) and the receiver (Y). The fourth image changes according to the condition: in the congruent condition, the receiver (Y) helps the communicator (X) by handing over his microphone, i.e. both CI and II are satisfied. In the semi-congruent condition, the observer (Z) helps the communicator by handing over her microphone, i.e. only II is satisfied. In the incongruent condition, the observer (Z) gives the communicator a deodorant, i.e., neither CI nor II are satisfied.



2.3 Procedure

Prior to the start of the experiment, participants were given a detailed description of the experimental methodology to be used and were required to sign an informed consent form. They were then positioned in front of a screen equipped with a sensor net for data collection and instructed to minimize movement. Stimuli were presented using E-prime 2.0 (Psychology Software Tools). The task procedure began with participants being instructed to look carefully at the visual stories that appeared on the screen. The visual stories were divided into 3 counterbalanced lists, A, B, and C. Participants assigned to list A viewed 90 visual stories, 30 congruent stories, 30 semi-congruent stories, 30 incongruent stories, and 15 fillers, for a total of 105 stimuli. Participants assigned to List B or List C received 90 visual stories containing the counterpart of each stimulus, plus 15 fillers.

The trial began with a fixation cross of 800 ms duration. Then the stimulus (visual story or filler) appeared on the screen. For the visual stories, each of the 4 pictures was presented for 4 s; for the fillers, the picture of the object appeared for 4 s. The trial ended with a gray screen displayed as an intertrial interval, randomly lasting 250/350 ms. A total of 90 trials for the visual stories and 15 fillers were randomly presented, for a total task duration of approximately 30 min. The experimental procedure is provided in Figure 2.

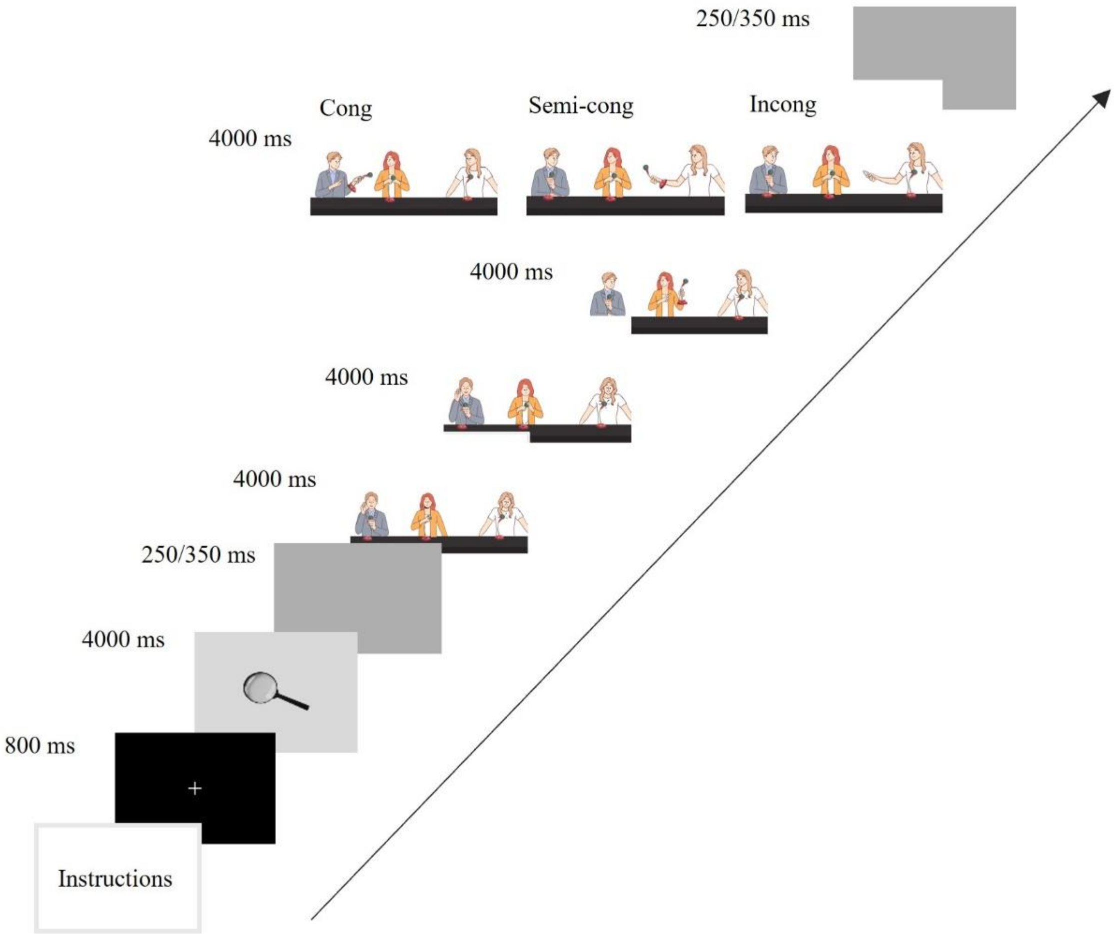


Figure 2. Experimental procedure.

2.4 Electroencephalographic data recording and processing

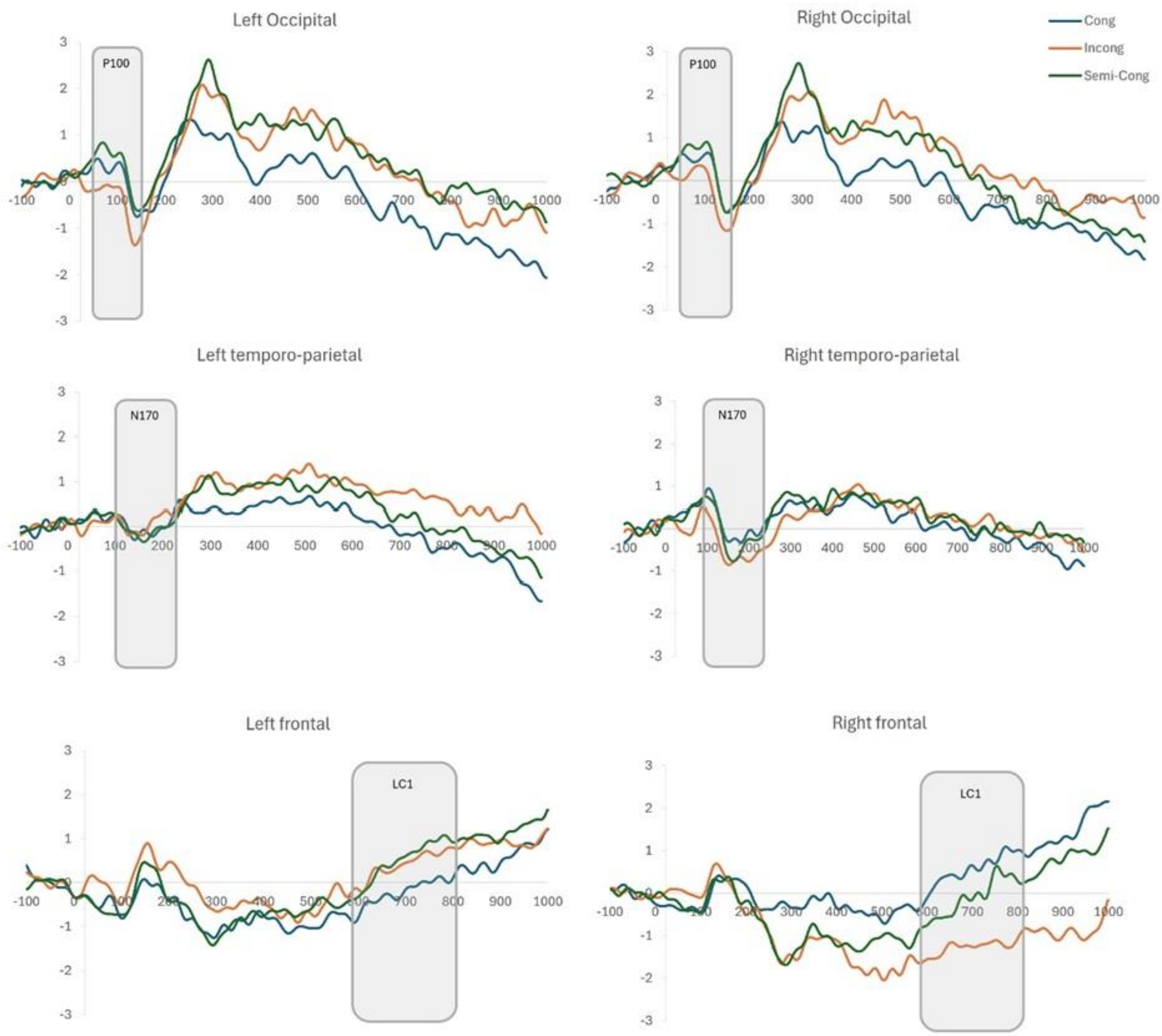
Electroencephalographic data were recorded continuously at 1000Hz sample rate using Net Station software (version 5.3.0.1, Electrical Geodesic, Inc., Eugene, OR, USA) and a 64-Hydrocel Geodesic Sensor Net, with impedances kept below 50 kΩ and the reference to the vertex (Cz). The acquired EEG data were processed using Net Station Tools (ibid.). The digital 30Hz low-pass filter was applied offline. The EEG data of each subject were segmented in epochs from 100 ms before the presentation of target picture to 1000 ms after the stimulus onset. The artefacts detection was set at 200 µV for bad channels, at 150 µV for the eye blinks and at 100 µV for the eye movements (Electrical Geodesic, Inc., Eugene, OR, USA; Anonymous 1 et al., 2022; Anonymous 2 et al., 2022; Anonymous 3 et al., 2013; McPartland et al., 2010). The segments with an eye blink, an eye movement or more than 30% bad channels were excluded. The data were re-referred to average of all the channels. Baseline correction -100 ms before the onset of the stimulus was applied.

Based on theoretical considerations and through visual inspection, we focused on extracting amplitude and latency of the following components and regions of interest (ROIs): P100 (50-150 ms) on occipital electrodes [left: 35(O1); right: 39(O2)]; P100-200 ms on frontal electrodes [left: 9(F1); 12(F3), 13(F5); right: 3,(F2), 60(F4), 59(F6)]; N170 (100-210 ms) on the temporo-parietal electrodes [left: 25(TP7), 29(TP9); right: 48(TP8), 47(TP10)]; P300 (250-400 ms) on centro-parietal electrodes [left: 21(CP1), 26(CP5); right: 41(CP2), 46(CP6)]; N350-600 on temporo-parietal electrodes. The mean amplitude of late components (LC1, 600-800 ms; LC2, 800-1000 ms) was extracted on frontal and temporo-parietal electrodes.

2.5 Statistical analysis

The repeated-measures ANOVAs with 3 Target (congruent vs. semi-congruent vs incongruent) *per* 2 Hemisphere (left vs. right) as within-subjects factors on amplitude and latency of ERPs components on each ROIs was performed. Statistical analyses were performed with Statistica v.7.0 (StatSoft Inc. 2004).

Figure 3. ERPs grand average on left and right occipital, temporo-parietal and frontal electrodes of the three conditions [Congruent (Cong), Incongruent (Incong), Semi-Congruent (Semi-Cong)]



3. Results

Analyses on the amplitude of the P100 (Figure 3) revealed a main effect of the Target at occipital electrodes [$F(2,84)= 4.04$; $p= .021$], with the congruent and the semi-congruent conditions showing a greater positive amplitude compared to the incongruent condition (respectively $p=.032$, $p= .009$)

The ANOVAs on the latency of the P100-200 on the frontal electrodes revealed an interaction effect of the Target *per* Hemisphere [$F(2,84)= 5.74$; $p= .005$]. Post-hoc analysis showed that on the right hemisphere the congruent condition had a greater latency compared to the incongruent condition ($p= .005$); in addition, the incongruent condition on the left hemisphere presented a greater latency compared to the incongruent condition on the right hemisphere ($p< 001$), and to the semi-congruent condition on the right hemisphere ($p= .036$); the semi-congruent condition on the left hemisphere showed a greater latency compared to the incongruent condition on the right ($p= .006$).

On the amplitude of N170 (Figure 3) at temporo-parietal electrodes, an interaction effect of Target *per* Hemisphere was found [$F(2,84)= 3.00$; $p= .054$]. Post-hoc analysis revealed that the incongruent condition on the right hemisphere had a greater negative amplitude compared to: the congruent condition on the same hemisphere ($p= .017$); the congruent condition on the left hemisphere ($p= .016$); the semi-congruent condition on the left hemisphere ($p= .036$); the incongruent condition on the left hemisphere ($p= .002$). Moreover, a main effect of the Hemisphere was found on the latency of the N170 [$F1,42= 14.73$; $p< 001$], with the right hemisphere showing a greater latency compared to the left one ($p< 001$).

Finally, the ANOVAs on the amplitude of the LC1 (600-800 ms) (Figure 3) on frontal electrodes revealed an interaction effect of Target *per* Hemisphere [$F(2, 84)=3,30$ $p= .042$], with the incongruent condition on the right hemisphere showing a lower amplitude compared to the right congruent condition ($p= .018$), and compared to the incongruent condition ($p= .020$) and to the semi-congruent condition ($p= .006$) on the left hemisphere.

The analysis did not show significant differences for the other selected components, such as P300 (250-400 ms) and LC2 (800-1000 ms).

The significant results of the amplitude and latency of the ERPs components are provided respectively in Table 1 and Table 2.

Table 1. Significant results of the 3 × 2 repeated-measures ANOVAs: Target [Cong vs. Semi-Cong vs. Incong] per Hemisphere [Left (L) vs. Tight (R)] on amplitude of the ERP components (P100, P100-200, N170, P300, N350-600, LC1, and LC2) on each region of interest (ROI) (occipital, parietal, frontal, temporo-parietal, and centre-parietal).

Component	ROI effect	Post-hoc
P100 (50-150 ms)	**Occipital** Target F(2,84)= 4.04; p= .021	Cong > Incong p= .032 Semi-Cong > Incong p= .009
N170 (100-210ms)	**Temporo-parietal** Target X Hemisphere F(2,84)= 3.00; p= .054	Incong R > Cong L p= .016 **Incong R > Cong R p= .017** Incong R > Incong L p= .002 Incong R > Semi-Cong L p= .036
N350-600 ms	**Temporo-parietal** /	
P300 (250-400 ms)	**Centre-parietal** /	
LC1 (600-800ms)	**Frontal** Target X Hemisphere F(2,84)= 3.30; p= .042	**Incong R < Cong R p= .018** Incong R < Incong L p= .020 Incong R < Semi-Cong L p= .006
LC2 (800-1000ms)	**Frontal** /	

Table 2. Significant results of the 3 × 2 repeated-measures ANOVAs: Target [Cong vs. Semi-Cong vs. Incong] per Hemisphere [Left (L) vs. Right (R)] on latency of the ERP components (P100, P100-200, N170, P300) on each region of interest (ROI) (occipital, frontal, temporo-parietal, and center-parietal).

Component	ROI effect	Post-hoc
P100 (50-150 ms)	**Occipital** /	
P100-200	**Frontal** Target X Hemisphere $F(2,84)= 5.74$; $p= .005$	**Cong R > Incong R p= .005** Incong L > Incong R p<001 Incong L > Semi-Cong R p=.036 Semi-Cong L > Incong R p=.006
N170 (100-210 ms)	**Temporo-parietal** Hemisphere $F(1,42)= 14.73$; $p <001$	R > L
P300 (250-400ms)	**Centre-parietal** /	

4. Discussion

This exploratory ERPs study investigated the neural correlates of processing of basic ostensive communicative acts, namely the time course of processing communicative and informative intentions expressed through a combination of eye contact and gestures. Participants observed visual stories with three agents: the first agent makes a request to a second agent while the third agent observes the scene. In making the request, the first agent manifests a communicative intention and an informative intention to the second agent. The stories had three possible outcomes, corresponding to three experimental conditions: a congruent condition, in which the second agent responds correctly to the first character's request (both intentions are satisfied); a semi-congruent condition, in which the third agent (not the second, as it should be) responds correctly to the first character's request (communicative intention is violated; only informative intention is satisfied); an incongruent condition, in which the third agent responds incorrectly to the first character's request (both communicative and informative intentions are violated). The stimuli were thus constructed to generate expectations about the successful outcome of the ostensive communicative act.

Three main findings emerged from the analysis of the ERPs. They related to the amplitude of two early components, i.e., P100 (50-150 ms) and N170 (100-210 ms), and one later component, i.e., LC1 (600-800 ms). In the following sections, we first discuss the specific results on the ERPs components in light of the existing ERPs literature. We then discuss the general implications of such results for the deflationist model of OC outlined in the introduction.

4.1.ERPs components

As for the first early component, the P100, the results showed that the incongruent condition differed from both the congruent and the semi-congruent conditions, which resulted in a greater positive amplitude at occipital electrodes. P100 wave is related to perceptual information processing in visual regions (Hillyard et al., 1995). Previous research show that its amplitude is associated with recruitment of attentional control (Dennis et al., 2007; Hillyard et al., 1998), for example is related to selective attention and the consumption of attention resource (Luck et al., 1994) and can be also affected by emotional facial expressions (Earls et al., 2016; Itier & Tylor, 2002; Moradi et al., 2017; Shah et al. 2018). From this view, the greater the amplitude of the P100, the more attentional control are recruited for stimulus processing. To exemplify the point: studies in patients with schizophrenia indicate that P100 amplitude is reduced in response to faces, indicating a decrease in visual attention to faces in these patients (Campanella et al., 2006; Earls et al., 2016; Murashko et al., 2019). With regard to the data of our study, the findings at P100 appear to indicate that, at the level of visual perceptual processing, the congruent and semi-congruent conditions imply an augmentation of early attentional resources. Specifically, congruent and semi-congruent stimuli appear to be processed in a similar manner and elicit a greater degree of attention in comparison to incongruent stimuli. Considering the evidence that P100 can influence subsequent stages of stimulus processing (Herrmann and Knight, 2001), we can hypothesize that the initial attentional capture in congruent and semi-congruent conditions may serve as a preparatory phase for activation of subsequent processes, such as those related to the second early component.

As for such component, the N170, an interaction effect of Target (congruent *vs*. semi-congruent *vs*. incongruent) by Hemisphere (right vs. left) at temporo-parietal electrodes was identified. The main post-hoc analysis concerns the right hemisphere and shows that the incongruent condition had a more pronounced negative amplitude compared to the congruent condition. Contrary to our predictions, there was no significant difference on this component between the semi-congruent and congruent conditions, nor between the semi-congruent and incongruent conditions. Previous literature suggests that the N170 exhibits larger amplitude (and shorter latency) to faces in comparison to other stimuli (e.g., objects) (Bentin et al., 1996). As such, it is presumed to reflect neural activity associated with early-stage face processing (Eimer, 2011; Hinojosa et al., 2015; Kang et al., 2018; Tautvydaitė et al., 2022). Specifically, prior studies show that the direction of gaze is discriminated during the N170. A larger and earlier N170 has been reported for gaze movements going away from, rather than back to the observer (Puce et al., 2000), suggesting the N170 is sensitive to the apparent motion of eyes. Some investigations report that the N170 is more negative in response to dynamic averted gaze shifts (Latinus et al., 2015; Rossi et al., 2015) or averted gaze face images (Itier et al., 2007; Watanabe et al., 2002) than to direct gaze counterparts. The results of our study are consistent with such findings. As shown in Figure 1, in the incongruent condition, the third agent delivers an (incorrect) object to the first agent, while the gaze of the first agent (who made the request to the second agent) is oriented in the opposite direction. The larger negative amplitude of the incongruent condition compared to the congruent condition thus indicates an early detection of such an inconsistency, i.e., the fact that the first and third agents are not looking at each other in the eyes (i.e., they are not communicating). This is further supported by latency analysis: at N170 in the right hemisphere, a significant difference was found between incongruent and congruent conditions. The absence of significant differences in the N170 response between congruent and semi-congruent conditions suggests that the processing observed in the semi-congruent condition, where only informative intention is satisfied, is analogous to that of the congruent condition, where both intentions are satisfied. This finding might be indicative that the processing of informative intention occurs at an early stage, approximately 170 milliseconds after the initial presentation, in a manner analogous to communicative intention. This finding thus might challenge the hypothesis of a two-step process for intention detection (CI) and content attribution (II).

It must be acknowledged that the N170 results do not preclude the possibility of subsequent processing of the two intentions. It is worth noting that the initial processing stage corresponds to processes related to perceptual mindreading (as we will discuss below), followed by the later stages of mind attribution. In this regard, although no significant differences were identified in the P300 component, as reported in Ruzzante and Vaes (2021), the impact of a more advanced stage of mind attribution appears to manifest later in the present study, within the 600-800 ms range, possibly due to the different nature of the experimental task. In this time window, at the right frontal electrodes, the incongruent condition showed a greater negative amplitude only compared to the congruent condition but not compared to the semi-congruent condition (in which the informative intention is satisfied). The frontomedial negativity (FMN), a family of event-related potential (ERP) deflections classically associated with performance monitoring (Holroyd and Coles, 2002), seems to be more pronounced when subjects receive an unfair versus a fair allocation (Aspé-Sánchez et al., 2020; Wu et al., 2011). These findings may be consistent with the results of the present study, in which the incongruent condition implies a greater cognitive effort for updating contextual information following the violation of both communicative and informative intentions.

4.2.Implications for a deflationist model of OC

The present study's findings appear to contribute to the OC deflationary hypothesis advanced by Moore (2016) and Sperber and Wilson (2024), adding a new piece to the puzzle. The results of this study show not only that basic forms of OC are indeed possible, but more importantly, they indicate that these basic forms are mentalistic in nature. That is, they seem to activate mindreading-related processes quite early. Crucial to this thesis are the results on the N170 related to eye and gaze processing, which indeed seem to indicate that the detection of both communicative and informative intentions takes place within the 200-millisecond window, thus indicating a low-level process of mindreading. As highlighted by many researchers (e.g., Emery, 2000; George and Conty, 2008; Itier and Batty, 2009; McCrackin and Itier, 2021), eye and gaze processing is at the core of social cognition. Attention to the eyes has been shown to be reduced in populations with altered mindreading, such as people with autism spectrum disorders and schizotypy, who indeed show atypicality on the N170 (Leung et al., 2021; Kang et al., 2018; O'Connor et al., 2007; Grice et al., 2005). Assuming that the N170 reflects an early activation of processes related to mindreading, which encompasses both types of intentions, it can be inferred that an early processing of content, corresponding to the level of informative intention, could occur much earlier than the mind attribution stage, which Ruzzante and Vause (2021) identify at 300 milliseconds.

If confirmed by future studies, such a finding might support the *direct perception model* of mindreading proposed by Gallagher (2008), a model that differs from both the TT and ST models discussed in the introduction. The first characteristic of perceptual mindreading concerns the way perception is understood. According to Gallagher, the perception one needs is "smart enough on its own, without the addition of inference mechanisms" (Gallagher, 2008, p. 536). In other words, the direct perception one needs must be simple enough to exclude extra-perceptual inference processes, but complex enough to capture the intentions and feelings of individuals. When I see my car in the parking lot below my house, the perception of what I see is due to complex brain processing. However, "the perception that I have of my car is direct – I see it right there in front of me. I do not have to glue anything together, add an interpretation or add an inference" (Gallagher, 2008, p. 537). Social perception, the perception of interactions that characterize the social relations of individuals, is no exception to the direct nature of perception: again, in "the usual circumstances of social interaction it does most of the work without the need of extra cognitive (theoretical or simulationist) processes" (Gallagher, 2008, p. 538). Such considerations lead Gallagher to argue that "In seeing the action and expressive movements of other person in the context of the surrounding world, one already sees their meaning; no inference to a hidden set of mental states (beliefs, desires, etc.) is necessary. When I see the other's action or gesture, I see (I immediately perceive) the meaning in the action or gesture. I see the joy, or I see the anger, or I see the intention in the face or in the posture or in the gesture or action of the other" (Gallagher, 2008, p. 542).

4.2.1 Implications for the development of ostensive communication

The considerations made so far point to the view that there are basic forms of OC, and that the processing of such basic forms relies on a type of processing that does not require reference to recursive metarepresentational systems, but is nevertheless a form of mindreading. This has important theoretical implications for both the ontogeny and phylogeny of OC. As mentioned above, the common intention of deflationary perspectives is to search for basic forms of OC that represent both an early stage of development of human communication in the ontogeny and a link to some forms of non-human communication. The experimental findings of the present study are consistent with those of previous research that has examined the case of human infants and non-human animals. These studies have significantly challenged the notion of the classical thesis, which posits that OC is exclusively a feature of human communication.

The ontogeny of human communication is one of the strongest lines of evidence highlighted by Sperber and Wilson (2024) in their revision of Relevance Theory. As they point out, "humans engage in ostensive communication, as both communicators and addressees, from the early age at which they start pointing communicatively: that is, well before they speak.” (Sperber and Wilson, 2024, p. 37) Informative-declarative pointing (i.e., pointing with the intention of providing new information to the recipient), which clearly involves both informative and communicative intentions, has indeed been demonstrated in human children from a very early age, as it reliably appears in infants around the age of 1 year (Butterworth and Morissette, 1996; Carpenter et al. 1998; Liszkowski et al., 2012). The robustness of pointing as a fully-fledged communicative device in prelinguistic children confirms that this form of ostensive communication does not depend on language, and if so, it likely does not require reference to recursive metarepresentational system of mindreading.

In contrast, declarative pointing, and especially informative-declarative pointing, is considered to be absent from the natural communicative repertoire of non-human apes (Tomasello, 2008). However, as already stressed by Gomez (1994) and Moore (2016, 2017), there is some evidence that basic forms of ostensive communication might already be present in non-human primates, especially in great apes. Basic ostensive communication mechanisms could help in cases where the communicative repertoire of a species is rather limited, as is the case for most nonhuman animals, including great apes (Sperber and Wilson, 2024). A first point highlighted by Gomez (1994) and Moore (2016) is the role of eye contact in great apes during interactions, especially in chimpanzees, a feature that is consistent with the role attributed to gaze in the present study. Indeed, eye contact has already been used in the past to label primate behaviors as instances of intentional communication, at least in studies focused on gestural communication in chimpanzees (Tomasello et al. 1985, 1989, 1994, 1997; Hobaiter and Byrne 2011, 2014; Roberts et al. 2012, 2013). Other examples come from Sperber and Wilson themselves (2024), such as the case of chimpanzees who make their communicative acts ostensive in mating requests, including the production of noisy signals such as leaf-clipping to attract attention when they want to mate (in the case of males, simultaneously showing the erect penis and shaking a branch); or, even more convincingly, as in the case of ostensive pre-mating persistence, which in some cases requires the production of various signals to make the request explicit and to convince the partner to enter into the mating relationship. As the two authors point out: "Repeating the request in various ways and modalities, showing persistence, patience, and flexibility elicits a greater expectation of relevance than a simple request would have done, and helps satisfy this expectation by giving evidence of what to expect from the requester’s behaviour in the consortship itself" (Sperber and Wilson, 2024, p. 31). As a final word of caution, however, we acknowledge that in non-human apes, it may be productive to distinguish between their cognitive capacity for ostensive communication and their social motivation to altruistically share honest information (cf. constraints on information donation, e.g. Burkart et al. 2018)

Before concluding, it is necessary to make some methodological considerations that may have influenced the results of the current research. As we have said, our data seem to indicate that basic forms of OC involve early processing of both informative and communicative intentions simultaneously. Moreover, our results seem to show that a perceptual mindreading system is involved in such a form of processing. Thus, both from the point of view of processing and from the point of view of the type of communication taking place, it is possible to speak of ostensive communication in the proper sense already in these basic forms. However, it is important to emphasize that our results are strongly related to the type of stimulus used. From this point of view, it is possible that the visual nature of the stimulus is the trigger condition for perceptual mindreading. It is therefore an open question to investigate whether linguistic communication also involves processes with similar characteristics to those highlighted in our study. It must be said, however, that the use of a stimulus that triggers perceptual processes is fully consistent with deflationist perspectives that aim to identify the basic conditions of OC from a phylogenetic perspective. Indeed, visual stimuli that activate perceptual processes are consistent with models of the pantomimic origin of language, i.e. models that claim that forms of communication based on (among other things) gestures, pantomime and facial expressions are the precursors of human language (e.g., Anonymous 4 et al., 2023; Arbib, 2024; Corballis, 2017; Anonymous 5 2022; Tomasello, 2008; Anonymous 6 et al., 2020). From this point of view, it is possible to hypothesize that a processing system such as perceptual mindreading is one of the basic conditions for the initiation of a properly human type of communication based on a gestural-visual expressive system.

5. Conclusion

The findings of the present study appear to indicate that the neural processing of the two intentions underlying ostensive communication—namely, communicative intention and informative intention—occurs in an early time frame, both within a window of 200 milliseconds. If substantiated by future research, these findings may support the hypothesis that basic forms of ostensive communication are governed by a form of perceptual mindreading. From a theoretical standpoint, these results challenge the classical theory of ostensive communication, which posits that the comprehension of the two intentions necessitates complex cognitive processes, often referred to as recursive mindreading. Moreover, such results allow us to further elucidate the revision of ostensive communication recently advanced by proponents of a deflationary hypothesis. Indeed, the data presented here, showing that communicative and informative intentions are processed together early, challenges the two-stage processing thesis that the deflationary hypothesis seems to imply. Overall, the idea of early (and therefore basic) processes underlying the elaboration of primitive forms of ostensive communication opens up much broader possibilities for future research on both its ontogeny and its phylogeny.

Acknowledgements

The authors thank all the participants who took part in the study.

Declarations

Data availability statements

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Conflict of interest

No potential conflict of interest was reported by the author(s).

Ethics approval statement

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The current study was approved by the Ethics Committee of “Masked Information for Blind Review”.

References

Anonymous 4 et al., 2023

Alcalá-López, D., Vogeley, K., Binkofski, F., & Bzdok, D. (2019). Building blocks of social cognition: Mirror, mentalize, share?. *Cortex*, *118*, 4-18. https://doi.org/10.1016/j.cortex.2018.05.006

Anonymous 1 et al., 2022

Arbib, M. A. (2024) Pantomime within and beyond the evolution of language, in P. Żywiczyński, J. Blomberg, M. Boruta-Żywiczyńska (eds), Perspectives on Pantomime, John Benjamins, Amsterdam, pp. 16-57.

Aspé-Sánchez, M., Mengotti, P., Rumiati, R., Rodríguez-Sickert, C., Ewer, J., & Billeke, P. (2020). Late frontal negativity discriminates outcomes and intentions in trust-repayment behavior. *Frontiers in psychology*, *11*, 532295. doi: 10.3389/fpsyg.2020.532295.

Bar-On, D. (2021). How to do things with nonwords: pragmatics, biosemantics, and origins of language in animal communication. *Biology & Philosophy*, *36*(6), 50. https://doi.org/10.1007/s10539-021-09824-z

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, *8*(6), 551-565.

Bohl, V., & van den Bos, W. (2012). Toward an integrative account of social cognition: marrying theory of mind and interactionism to study the interplay of Type 1 and Type 2 processes. *Frontiers in Human Neuroscience, 6*, 274.

Brink, I. (2004). Joint attention, triangulation and radical interpretation: A problem and its solution. *Dialectica*, *58*(2), 179-206. https://doi.org/10.1111/j.1746-8361.2004.tb00296.x

Burkart, J., Guerreiro Martins, E., Miss, F., & Zürcher, Y. (2018). From sharing food to sharing information: cooperative breeding and language evolution. Interaction Studies, 19(1-2), 136-150

Butterworth, G., & Morissette, P. (1996). Onset of pointing and the acquisition of language in infancy. *Journal of reproductive and infant psychology*, *14*(3), 219-231. https://doi.org/10.1080/02646839608404519

Anonymous 3 et al., 2013

Campanella, S., Montedoro, C., Streel, E., Verbanck, P., & Rosier, V. (2006). Early visual components (P100, N170) are disrupted in chronic schizophrenic patients: an event-related potentials study. *Neurophysiologie Clinique/Clinical Neurophysiology*, *36*(2), 71-78.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, i-174. https://doi.org/10.2307/1166214

Anonymous 2 et al., 2022

Corballis, M. C. (2017). *The truth about language: what it is and where it came from*. University of Chicago Press, Chicago.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language, 25*(2), 141–168. https://doi.org/10.1111/j.1468-0017.2009.01384.x

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.

Davies, M., & Stone, T. (eds) (1995a). *Folk Psychology*, Blackwell.

Davies, M.. & Stone, T. (eds) (1995b) Mental Simulation, Blackwell

Dennis, T. A., & Chen, C. C. (2007). Neurophysiological mechanisms in the emotional modulation of attention: the interplay between threat sensitivity and attentional control. *Biological psychology*, *76*(1-2), 1-10.

Earls, H. A., Curran, T., & Mittal, V. (2016). Deficits in early stages of face processing in schizophrenia: a systematic review of the P100 component. *Schizophrenia bulletin*, *42*(2), 519-527.

Eimer, M. (2011). The face-sensitivity of the n170 component. *Frontiers in human neuroscience*, *5*, 119.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews*, *24*(6), 581-604. https://doi.org/10.1016/S0149-7634(00)00025-7

Anonymous 5 2022

Fodor, J. (1987). *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*. MIT Press, Cambridge.

Gallese, V. (2013). Mirror neurons, embodied simulation and a second-person approach to mindreading. *Cortex*, *49*(10), 2954-2956.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493-501.

Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation?. *Trends in cognitive sciences*, *15*(11), 512-519. https://doi.org/10.1016/j.tics.2011.09.003

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and cognition*, *17*(2), 535-543. https://doi.org/10.1016/j.concog.2008.03.003

George, N., & Conty, L. (2008). Facing the gaze of others. *Neurophysiologie Clinique/Clinical Neurophysiology*, *38*(3), 197-207. https://doi.org/10.1016/j.neucli.2008.03.001

Goldman, A.I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, New York.

Gómez, J. C. (1994). Mutual awareness in primate communication: A Gricean approach. In S.T. Parker, R. Mitchell, & M. Boccia (Eds.), *Self-Awareness in animals and humans*. Cambridge University Press.

Gómez, J. C. (1996). Ostensive behavior in great apes. The role of eye contact. In A.E. Russon, K.A. Bard, & S.T. Parker (Eds.), Reaching into thought: The minds of the great apes (pp. 131–151). Cambridge University Press.

Grice, H.P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press

Grice, H. P. (1957). Meaning. *The philosophical review*, *66*(3), 377-388.

Grice, S. J., Halit, H., Farroni, T., Baron-Cohen, S., Bolton, P., & Johnson, M. H. (2005). Neural correlates of eye-gaze detection in young children with autism. *Cortex*, *41*(3), 342-353.

Herrmann, C. S., & Knight, R. T. (2001). Mechanisms of human attention: event-related potentials and oscillations. *Neuroscience & Biobehavioral Reviews*, *25*(6), 465-476.

Hillyard S. A., Mangun G. R., Woldorff M. G., Luck S. J. (1995). Neural mechanisms mediating selective attention. In Gazzaniga M. S. (Ed.), The Cognitive Neurosciences (pp. 320–67). Cambridge: MIT Press.

Hillyard S. A., Vogel E. K., Luck S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electro-physiological and neuroimaging evidence. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 353(1373), 1257–270. doi: 10.1098/rstb.1998.0281

Hinojosa, J. A., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *55*, 498-509.

Hobaiter, C., & Byrne, R. W. (2011). The gestural repertoire of the wild chimpanzee. *Animal cognition*, *14*, 745-767. https://doi.org/10.1007/s10071-011-0409-2

Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Current Biology*, *24*(14), 1596-1600.

Holroyd, C. B., and Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709. doi: 10.1037/0033-295x.109.4.679

Hostetter, A. B., Cantero, M., & Hopkins, W. D. (2001). Differential use of vocal and gestural communication by chimpanzees (Pan troglodytes) in response to the attentional status of a human (Homo sapiens). *Journal of Comparative Psychology*, *115*(4), 337.

Kang, E., Keifer, C. M., Levy, E. J., Foss-Feig, J. H., McPartland, J. C., & Lerner, M. D. (2018). Atypicality of the N170 event-related potential in autism spectrum disorder: a meta-analysis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(8), 657-666.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in cognitive sciences*, *11*(5), 194-196. https://doi.org/10.1016/j.tics.2007.02.002

Itier, R. J., Alain, C., Kovacevic, N., & McIntosh, A. R. (2007). Explicit versus implicit gaze processing assessed by ERPs. *Brain research*, *1177*, 79-89.

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neuroscience & Biobehavioral Reviews*, *33*(6), 843-863. https://doi.org/10.1016/j.neubiorev.2009.02.004

Itier R. J., Taylor M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. Neuroimage, 15(2), 353–72. doi: 10.1006/nimg.2001.0982

Latinus, M., Love, S. A., Rossi, A., Parada, F. J., Huang, L., Conty, L., ... & Puce, A. (2015). Social decisions affect neural activity to perceived dynamic gaze. *Social cognitive and affective neuroscience*, *10*(11), 1557-1567.

Leung, C., Lei, K. S., Wang, S. M., & Lam, B. Y. H. (2021). Theory of mind in schizotypy: a behavioral and event-related potential (ERP) study. *Schizophrenia Research: Cognition*, *23*, 100190. https://doi.org/10.1016/j.scog.2020.100190

Liebal, K., Pika, S., Call, J., & Tomasello, M. (2004). To move or not to move: how apes alter the attentional state of humans when begging for food. *Interaction Studies*, *5*, 199-219.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & De Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive science*, *36*(4), 698-713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732-739. https://doi.org/10.1016/j.cognition.2008.06.013

Loria, E. (2017). The role of the ostensive communicative context in the childhood social learning. In *Modeling and Using Context: 10th International and Interdisciplinary Conference, CONTEXT 2017, Paris, France, June 20-23, 2017, Proceedings 10* (pp. 315-323). Springer International Publishing.

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: psychophysical and electrophysiological evidence for early selection. *Journal of experimental psychology: human perception and performance*, *20*(4), 887.

McCrackin, S. D., & Itier, R. J. (2021). Feeling through another's eyes: Perceived gaze direction impacts ERP and behavioural measures of positive and negative affective empathy. *NeuroImage*, *226*, 117605. https://doi.org/10.1016/j.neuroimage.2020.117605

McPartland, J., Cheung, C. H., Perszyk, D., & Mayes, L. C. (2010). Face-related ERPs are modulated by point of gaze. *Neuropsychologia*, *48*(12), 3657–3660. https://doi.org/10.1016/j.neuropsychologia.2010.07.020

Moore, R. (in press), Intentions in human and non-human great ape communication. In B. Geurts and R. Moore (Eds.), *Evolutionary pragmatics*. Oxford: Oxford University Press.

Moore, R. (2016). Meaning and ostension in great ape gestural communication. *Animal Cognition*, *19*(1), 223-231.

Moore, R. (2017). Gricean communication and cognitive development. *The Philosophical Quarterly*, *67*(267), 303-326.

Moradi, A., Mehrinejad, S. A., Ghadiri, M., & Rezaei, F. (2017). Event-related potentials of bottom-up and top-down processing of emotional faces. *Basic and clinical neuroscience*, *8*(1), 27.

Murashko, A. A., & Shmukler, A. (2019). EEG correlates of face recognition in patients with schizophrenia spectrum disorders: A systematic review. *Clinical neurophysiology*, *130*(6), 986-996.

O'Connor, K., Hamm, J. P., & Kirk, I. J. (2007). Neurophysiological responses to face, facial regions and objects in adults with Asperger's syndrome: an ERP investigation. *International Journal of Psychophysiology*, *63*(3), 283-293.

Origgi, G., & Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers and A. Chamberlain (eds), Evolution and the Human Mind: Modularity, Language and Meta-Cognition. Cambridge: Cambridge University Press.

Povinelli, D. J., Theall, L. A., Reaux, J. E., & Dunphy-Lelii, S. (2003). Chimpanzees spontaneously alter the location of their gestures to match the attentional orientation of others. *Animal behaviour*, *66*(1), 71-79.

Puce, A., Smith, A., & Allison, T. (2000). ERPs evoked by viewing facial movements. *Cognitive neuropsychology*, *17*(1-3), 221-239. https://doi.org/10.1080/026432900380580

Roberts, A. I., Vick, S. J., & Buchanan-Smith, H. M. (2012). Usage and comprehension of manual gestures in wild chimpanzees. *Animal Behaviour*, *84*(2), 459-470.

Roberts, A. I., Vick, S. J., & Buchanan-Smith, H. M. (2013). Communicative intentions in wild chimpanzees: persistence and elaboration in gestural signalling. *Animal Cognition*, *16*, 187-196.

Rossi, A., Parada, F. J., Latinus, M., & Puce, A. (2015). Photographic but not line-drawn faces show early perceptual neural sensitivity to eye gaze direction. *Frontiers in Human Neuroscience*, *9*, 185.

Ruzzante, D., & Vaes, J. (2021). The timeline of mentalization: Distinguishing a two-phase process from mind detection to mind attribution. *Neuropsychologia*, *160*, 107983. https://doi.org/10.1016/j.neuropsychologia.2021.107983

Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural correlates of mental state decoding in human adults: an event-related potential study. *Journal of cognitive neuroscience*, *16*(3), 415-426.

Scott-Phillips, T. (2015a). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Bloomsbury Publishing.

Scott-Phillips, T. C. (2015b). Meaning in animal and human communication. Animal Cognition, 18(3), 801-805. https://doi.org/10.1007/s10071-015-0845-5

Scott-Phillips, T., & Heintz, C. (2023). Great ape interaction: Ladyginian but not Gricean. *Proceedings of the National Academy of Sciences*, *120*(42), e2300243120. https://doi.org/10.1073/pnas.2300243120

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current biology*, *18*(9), 668-671.

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition*, *108*(2), 303-319. https://doi.org/10.1016/j.cognition.2008.02.009

Shannon, C.E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana.

Schulze, C., & Tomasello, M. (2015). 18-month-olds comprehend indirect communicative acts. *Cognition*, *136*, 91-98. https://doi.org/10.1016/j.cognition.2014.11.036

Shah, D., Knott, V., Baddeley, A., Bowers, H., Wright, N., Labelle, A., ... & Collin, C. (2018). Impairments of emotional face processing in schizophrenia patients: Evidence from P100, N170 and P300 ERP components in a sample of auditory hallucinators. *International Journal of Psychophysiology*, *134*, 120-134.

Siposova, B., Tomasello, M., & Carpenter, M. (2018). Communicative eye contact signals a commitment to cooperate for young children. *Cognition*, *179*, 192-201.

Sperber, D. (2000) Metarepresentations in an evolutionary perspective. In D. Sperber (ed.) *Metarepresentations: An Interdisciplinary Perspective*. New York: Oxford University Press.

Sperber, D., & Origgi, G. (2010). A pragmatic account of the origin of language. In: Larson, R.K., Déprez, V., Yamakido, H. (Eds.), The Evolution of Human Language: Biolinguistic perspectives. Cambridge University Press, Cambridge, pp. 124–132.

Sperber, D., & Wilson, D. (2024). Rethinking ostensive communication in an evolutionary, comparative, and developmental perspective. *PsyArXiv: https://doi. org/10.31234/osf. io/zp3fx*.

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and cognition. Second Edition*. Blackwell, Oxford UK.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language*, 17 (1–2), 3–23. https://doi.org/10.1111/1468-0017.00186

Tautvydaitė, D., Mares, I., Rahman, M. S., Burra, N., & Senju, A. (2022). Effect of perceived eye gaze on the N170 component–A systematic review. *Neuroscience & Biobehavioral Reviews*, *143*, 104913.

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press, Cambridge.

Tomasello, M., George, B. L., Kruger, A. C., Jeffrey, M., & Evans, A. (1985). The development of gestural communication in young chimpanzees. *Journal of Human Evolution*, *14*(2), 175-186.

Tomasello, M., Gust, D., & Frost, G. T. (1989). A longitudinal investigation of gestural communication in young chimpanzees. *Primates*, *30*, 35-50.

Tomasello, M., Call, J., Nagell, K., Olguin, R., & Carpenter, M. (1994). The learning and use of gestural signals by young chimpanzees: A trans-generational study. *Primates*, *35*, 137-154.

Tomasello, M., Call, J., Warren, J., Frost, G. T., Carpenter, M., & Nagell, K. (1997). The ontogeny of chimpanzee gestural signals: a comparison across groups and generations. *Evolution of communication*, *1*(2), 223-259.

Watanabe, S., Miki, K., & Kakigi, R. (2002). Gaze direction affects face perception in humans. *Neuroscience letters*, *325*(3), 163-166.

Wu, Y., Leliveld, M. C., and Zhou, X. (2011). Social distance modulates recipient's fairness consideration in the dictator game: an ERP study. *Biol. Psychol.* 88, 253–262. doi: 10.1016/j.biopsycho.2011.08.009

Yousefi Heris, A. (2020). How Might Simulation-Based Accounts of Mindreading Explain Pragmatic Interpretation?. *Phenomenology and Mind*, (19), 226-240.

Anonymous 6 et al., 2020

4. Summary and Conclusion

In this work, I showed the role of low-level mechanisms in ostensive communication, highlighting the contribution of motor simulation in language acquisition during development; I presented a review of experimental semiotics studies, indicating how to examine the foundations of pragmatic competence through the analysis of the way in which a communication system is established from scratch; and I have conducted an EEG study on ostensive communication, showing that there is a basic form of it that does not require the involvement of high-level mindreading. This work started with explaining what ostensive communication is in its classical formulation, starting from Grice's framework through Sperber & Wilson's relevance theory, according to whom (in their initial conception) ostensive communication is an exclusively human activity, shaped by the expression and recognition of communicative and informative intentions. In this way, ostensive communication is framed as a high-level cognitive process, requiring metarepresentational abilities in order to represent others' mental states in a recursive way. One of the problems of this approach is that it characterizes ostensive communication as an all-or-nothing phenomenon, making it difficult to place it in an evolutionary framework, and making an apparently unbridgeable gap between animal communication and human communication (language first and foremost).

However, in recent years a deflationary approach to this model has emerged, trying to fill the gap by arguing that basic forms of ostensive communication exist in human infants and non-human primates, re-evaluating the underlying cognitive mechanisms, i.e. the type of mindreading implicated in this basic ostensive communication. While classical perspectives argue for high-level inferential mindreading in the processing of communicative and informative intentions, deflationary perspectives link more basic forms of mindreading to correspondingly basic forms of ostension.

This dissertation aimed to contribute to resolving the debate between these competing models by investigating the processing dynamics of communicative and informative intentions conveyed through eye contact and gestures. To this end, we performed a study involving basic forms of ostensive communication, and identified the neural mechanisms at work in the recognition of these intentions, investigating whether they are high- or low-level processes. The results of our ERP analysis support a deflationary perspective in relation to basic forms of ostensive communication, whereby both communicative and informative intentions are detected within the first 200 ms from exposure to the stimulus, consistent with low-level mindreading. However, it is necessary to specify that those analyzed here are basic forms of ostensive communication: this means that **our study provided evidence in favor of the existence of basic ostensive communication that - since it does not require elaborate forms of mindreading - could be part of the communicative repertoire of children who do not yet possess a full-fledged ToM** (thus laying the foundations for a description of the development of pragmatic skills in humans), **and of non-human primates, specifically apes**.

This work is also compatible with a research path in cognitive sciences that tries to reevaluate embodied processes, so that instead of reducing complex cognitive phenomena to the latter, it seeks to ground them in those processes. In this sense, experimental semiotics provides an important tool for exploring the origins and development of pragmatic skills, including ostensive communication. Experimental semiotics is perfectly suited to the study of how communicative conventions emerge and provides the possibility of deepening the role of the related cognitive mechanisms involved. In the union of semiotic theory with cognitive sciences, experimental semiotics offers a tool for anchoring the study of communication to its evolutionary history, making possible a deeper understanding of the origin and evolution of language. An important feature of experimental semiotics studies is that **without forms of cooperation** between participants, **it is difficult or impossible to bootstrap and maintain a communication system, when trying to build it from scratch**.

A key point is that, if we look at ostensive communication and pragmatic competence as examined in this thesis, it turns out that there are basic cognitive abilities that probably come before syntactic and semantic competence. An example that has been brought is that of prelinguistic children: even before speaking, humans demonstrate their capacity for ostensive communication, acting as both senders and receivers of intentional signals, beginning with communicative pointing. It is these basic pragmatic skills that provide support for language development. However, as we have seen, we showed that basic pragmatic skills could also be present in non-human primates. An important conclusion is that **the recognition of communicative intentions alone is not sufficient to explain the distinguishing aspects of natural language, and therefore cannot explain by itself the origin of language**. To do this, we need to investigate which specific aspects of pragmatic competence played a fundamental role in the origin of natural language. If, as we have seen, ostensive communication is anchored, in development, to low-level mechanisms such as motor simulation, then the expression and recognition of communicative intentions can potentially be shared by other species besides humans. And this is indeed the result of our EEG study on mindreading and ostensive communication. However, as said, experimental semiotics shows us that cooperation is important for the ability to communicate typical of our species. This means that **the specific human modes of cooperation could be anchored in cognitive mechanisms that go beyond the mere fact of being ostensive communicators**. In fact, either we are "special" ostensive communicators (in the sense that our ostensive abilities are more extensive and powerful than those of other species), or the foundations of symbolic communication and natural language are to be found in other abilities. This is a crucial difference from the theories on the uniqueness of ostensive communication, and on its role for the origin of language supported by other authors, as we have seen throughout the thesis.

Future challenges should be focused on the questions that have not been answered here but that this work suggests, i.e. what makes human communication what it is, that is, how to explain the typical properties of natural language and its differences with animal communication that were discussed in the introduction. We also mentioned that these specificities should be explained by a set or core of cognitive abilities that could be unique to humans, in the first hypothesis, or differ in degree, in the second one. Having brought here evidence for the existence of basic forms of ostensive communication, one possibility is, again, that the differences are not due to this specific aspect of human pragmatic competence. Another answer could instead be that our ostensive communication differs in degree or extent of its domain from the basic forms of ostensive communication. In order to understand this, it is crucial to focus research on the ecological aspects that had an impact, from an evolutionary point of view, on human communication. For example, one approach could be the one linked to the investigation of epistemic vigilance processes that favored the emergence and stabilization of cognitive abilities suited to our modes of communication. From this perspective, research focused on pragmatic aspects of communication - experimental semiotics included - could provide us with some additional elements to support or discredit this or other hypotheses.

5. Declarations by all co-authors

Individual contribution to "Experimental Semiotics: A Systematic Categorization of Experimental Studies on the Bootstrapping of Communication Systems"

Angelo Delliponti contributed to conceiving and designing the analysis, collecting and coding the data, conceptualizing and writing the manuscript, and doing the cluster analysis. Renato Raia contributed to conceiving and designing the analysis, collecting and coding the data, and writing the manuscript. Giulia Sanguedolce contributed to conceiving and designing the analysis, collecting and coding the data, and conceptualizing and writing the manuscript. Adam Gutowski contributed to conceiving and designing the analysis, collecting and coding the data, and conceptualizing and writing the manuscript, did the cluster analysis, and provided figures. Michael Pleyer contributed to conceiving and designing the analysis, collecting and coding the data, and conceptualizing and writing the manuscript. Marta Sibierska contributed to the coding and to conceptualizing and writing the manuscript. Marek Placiński contributed to the coding and to conceptualizing and writing the manuscript, did the cluster analysis, and prepared the figures. Przemysław Żywiczyński contributed to the coding and to conceptualizing and writing the manuscript. Sławomir Wacewicz contributed to conceiving and designing the analysis, collecting and coding the data, and conceptualizing and writing the manuscript. All authors reviewed the manuscript.

Individual contribution to "Which mindreading for ostensive communication? An event-related potentials study of how the brain processes communicative and informative intentions"

Francesco Ferretti planned the study, supervised the recruitment of participants and administration of the task, interpreted the results, and wrote the paper. Angelo Delliponti planned the study, contributed to the preparation of the experimental materials, recruited the participants, administered the task, contributed to the interpretation of the results, contributed to the final draft. Valentina Deriu contributed to the design of the study, to the preparation of the experimental materials, processed the data, contributed to the interpretation of the results and to the final draft. Alessandra Chiera contributed to the design of the study, to the preparation of the experimental materials, to the interpretation of the results and to the final draft. Daniela Altavilla contributed to the design of the study, to the processing of the data, to the interpretation of the results and to the final draft. Serena Nicchiarelli drew the figures and contributed to the interpretation of the results. Sławomir Wacewicz contributed to the design of the study, to the interpretation of the results, and to the final draft. Ines Adornetti contributed to the design of the study, supervised the recruitment of participants and administration of the task, contributed to the interpretation of the results and to the final draft.