Theory of Mind in non-linguistic animals: a multimodal approach

Ph.D candidate
Luca Di Vincenzo

First supervisor
Simone Flaviano Pollo

Second supervisor
Marina De Palo

SAPIENZA
UNIVERSITÀ DI ROMA

SAPIENZA
UNIVERSITÀ DI ROMA

Theory of Mind in non-linguistic animals: a multimodal approach

Faculty of Arts and Humanities
Philosophy Department
Ph.D. in Philosophy

Luca Di Vincenzo
Identification no. 1637053

First supervisor	Second supervisor
Simone Flaviano Pollo	Marina De Palo

A.Y. 2023/2024

“So long and thanks for all the fish [… again and again!]”

(Douglas Adams)

Acknowledgements

I am not used to writing acknowledgments. But this time I will make an exception. I would like to thank all those who have accompanied me during these three years of research: my supervisors Simone Pollo and Marina De Palo, for showing me which paths to follow and which to avoid, and for giving me trust; Robert Lurz, for teaching me a method of organizing concepts and thoughts; Tito Magri, for the experience he gave me; Antonella Mallozzi, for encouraging me to write; Massimiliano Di Giovanni, for his enthusiasm and willingness to explore the topics of my research; Vesta Eleuteri, for helping with the experimental part of the research; My colleagues Edoardo Morè, Francesco Di Raimondo, Morgana Bizzego and Filippo Sanna, because without you the trip would not have had the same meaning; José Luis Bermùdez for initiating me into the study of the philosophy of the animal mind and cognitive science; Cody Cedergren for welcoming me into his family; Enrique Chávez-Arvizo for his insights during the seminars organized in New York City; Daniel Campos, for his peripatetic walks; Tommaso Lamarra for listening to my doubts and trying to clarify them; Alessio Cutrone for always being there and why he will always be there; my roommates Davide Di Terlizzi, Samuele Cerquetelli and Musetta Paglialunga for making me feel at home; Leonardo Vannucci for his endless and never to be taken for granted support, even in the most difficult moments; Valentina Calà for making me grow; Eliana Streppa for believing in me; Emanuele Zecchino for allowing me to be part of his story; Susanna Tranelli for always teaching me the meaning of the word friendship; Andrea Erdogan Bonifazio for always teaching me to tread lightly; my family, especially my father Paolo Di Vincenzo and my mother Carla Milone[1] for never standing in my way and my sister Savina Di Vincenzo for never making me feel lonely; Roberto De Pau for being a beacon; Alessandra D'Onofrio for the encouragement and for always being an example to me.

Finally, because they are the most important, I would like to thank the non-human animals who have accompanied me so far and all those who will accompany me in the future. Among them, a special thanks to my dogs and my turtle: Titta, Lucky, Nahla, Rugantico. You taught me everything.

Thank you.

---

[1] Special thanks also for hand drawing, on silk, all the illustrations shown in this paper.

Contents

Preface

The debate surrounding animal cognition has raised fundamental questions that cross disciplinary boundaries in philosophy, psychology, and neuroscience. One of the key issues in this debate is the ability of nonhuman animals to perceive, recognize, and potentially attribute mental states to other beings, whether conspecific or interspecific. The question of whether non-human animals possess cognitive abilities has been historically approached with a focus on language. The central inquiry is whether an animal without language can engage in thought. However, this view has been challenged, and it is now recognized that animals without language can engage in complex cognitive processes. This perspective assumes that thought and language are strictly dependent on each other. In recent years, cognitive science's interdisciplinary approach has resolved the question of whether animals without language can think positively. Various nonhuman animals behave in a way that necessarily requires recognition of their capacity for thought or reasoning. The question now is what the limits of this thinking capacity are. The question arose as to whether these limits would prevent one from ascribing the capacity for thought not only to oneself but also to others. While the ability to think has been recently freed from the need for language, the same cannot be said for the ability to interpret the behaviour of others as caused by thought, which is still argued to be linked to linguistic abilities.

This research is based on the philosophical theory of representationalism, which posits that knowledge of the external world is derived from internal mental representations. The mind processes information by perceiving and interpreting its surroundings through internal constructs, rather than through direct contact with external stimuli. According to the representational hypothesis of the mind, one must be able to meta-represent in order to interpret the behaviour of others as caused by a thought. To illustrate this concept, consider the scenario of two friends, Alice and Bob, planning to go to the movies. Alice suggests seeing movie A, unaware that Bob has already seen it. One possible example of meta-representation is when Alice assumes that Bob has not seen movie A, but Bob is aware that he has already seen it. When Alice proposes watching movie A, Bob realizes that she is unaware of this information and is therefore meta-representing. Bob has a representation of Alice's representation, thinking that she thinks he has not seen the movie.

Returning to the representational hypothesis of the mind, in order to meta-represent, one must have the conceptual ability to articulate representations in propositional form. Therefore, to be able to think thoughts, whether they are our own or others, language is necessary. The necessity of language arises from its hierarchical organization of systems of symbols, which allows for the application of recursive and logical rules to representations. Clear relationships between thoughts are essential for their formation. Comparative studies between language and other communicative systems highlight that nonhuman animal communication systems are organized sequentially rather than hierarchically, which does not provide the tools to construct hierarchical relationships between thoughts. In order to determine whether animals other than humans are capable of specific cognitive abilities, researchers have studied the unique features of their communication systems. The studies resulted in identifying recursiveness as a characteristic unique to humans. This led to reformulating the initial assertion that meta-representation was impossible without language to state that it is impossible without recursiveness to represent the thoughts of others.

Against this theoretical background, this research aims to highlight an important bias that characterizes the analysis of animal communicative systems: unimodality. Unimodality refers to the exclusive analysis of the sound component of communication. Communication systems of nonhuman animals, as well as human language, have always been studied by placing comparative emphasis on the component that prevails in communicative exchanges between humans: voice. However, who determines that a hierarchically structured symbol system must be present within a single mode of expression and not perhaps emerge from the integration of multiple modes, if engaging in meta-representations requires it?

The research aimed to explore the relationship between meta-representations and language in animals. Specifically, it sought to demonstrate that certain linguistic features previously thought to be unique to humans may also be present in other animal communication systems. Collecting a large amount of behavioural data on different species, collaborating with different research centers, and using machine-learning techniques of automatic classification would have been necessary to accomplish this task. The goal was to detect the presence of hierarchically structured symbol systems across all communicative modalities employed by a given species, not just within a single modality. Contact had been initiated with the Primate Research Institute in Kyoto for this purpose.

Due to the COVID-19 pandemic lockdown, the initially proposed goals could not be implemented, resulting in a reformulation of the objectives. The original aim was to explore the epistemological question of whether non-linguistic animals could engage in meta-representations. What has not been previously mentioned is that even when research has successfully demonstrated the presence of recursiveness in other forms of communication and revised the conceptual framework with which we interpret the nature of the mind today, this investigation would still remain an a priori inquiry. In the field of cognitive science, when investigating whether non-linguistic animals can have meta-representations, two questions arise. One is metaphysical, concerning the nature of the mind and often conditioning scientific research. The other is epistemological, concerning the building of experimental paradigms suitable for demonstrating the presence or absence of a certain cognitive capacity in a given species. Additionally, it is important to note that the metaphysical inquiry is always secondary to the epistemological one. Demonstrating a fact empirically and without controversy conditions the theoretical framework and necessitates a thorough revision. Unable to deal with its original aim, this research focused on the epistemological question of designing experimental protocols capable of detecting the presence of meta-representational abilities in non-linguistic animals.

In the first chapter, we examine the history, importance, and biases of research on theory of mind (ToM). ToM refers to the ability to represent the mental states of others, including non-linguistic animals. Two things stand out in particular. Firstly, the bias of the multimodal approach to the study of the communicative systems of non-linguistic animals, already mentioned in this introduction; and secondly, that in the current state of research, of all the experimental protocols tested to detect the presence of ToM in non-linguistic animals, none has been able to provide evidence that is not controversial. However, investigating the multimodal nature of the minds and communicative systems of non-linguistic animals can reveal new behavioural phenomena that can be exploited to design more effective experimental designs for detecting ToM in non-linguistic animals.

In the second chapter, this was done, first by reconstructing the history of studies of language and other communicative systems in both humans and nonhuman animals, and then by reviewing the studies of multimodality that have emerged especially since the 1970s, both in humans and especially in nonhuman animals. Two phenomena emerged from this literature review that are relevant to the research purpose: the first is the multimodal shift, which refers to the ability to use the most appropriate communication channel for a given context. The second phenomenon is the correlation between ToM and language development in humans and the degree of development of cross-modal binding, the ability to integrate and harmonize information from different sensory inputs.

In the third chapter, a theoretical basis for the experimental protocols that are to be designed is provided. The development of a theory on the nature of mind is based on multimodal shift and cross-modal binding, which is referred to as multimodal mind-theory. Multimodal mind theory aims to decouple language and ToM, thus addressing both the metaphysical and the epistemological problem. Based on two proposed thought experiments, multimodal mind theory suggests that language and other communication systems, as well as ToM, are not in an interdependent relationship, but depend on the degree of development of cross-modal binding.

The fourth chapter addresses the epistemological question, focusing on experimental protocols proposed by Lurz that employ illusory environments and novel tasks for the animal being tested. These protocols are deemed adequate for investigating the presence of ToM in non-linguistic animals. A new protocol is developed from these existing protocols, utilizing the multimodal shift mechanism to enhance functionality. The first part of this protocol is practically applied in a pilot study involving two Asian elephants at the Rome Zoo. The study suggests that the elephants are capable of attributing perceptual states to other individuals.

By assuming that nonhuman animals can engage in ToM attitudes, the research does not address the main metaphysical theories that deny this possibility a priori. Using multimodal mind-theory as a theoretical background, the fifth chapter analyzes the two main theses against the possibility of ToM in nonhuman animals: Davidson's and Bermùdez'. By reconstructing the debate around their theses and analyzing their critical issues, it is demonstrated that multimodal mind-theory is better suited to explain the nature of mind than their proposals.

1 Significance, History, and Gaps of the Theory of Mind in Non-Linguistic Animals

When observing another individual, whether human or animal, it is natural to wonder about their mental state. This is where the concept of Theory of Mind (ToM) comes into play. ToM refers to the ability to attribute mental states, such as beliefs, desires, emotions, intentions, perceptions, and knowledge to others, in order to predict and interpret their behaviour. ToM[1], along with language, is recognized as one of the most sophisticated cognitive abilities of human beings. It is used in various everyday activities, such as attributing mental states to others (explicitly or implicitly) when asking if they are worried or sad and need to talk. It is also used to cooperate or compete, such as anticipating the opponent's moves in a sports competition by inferring their intentions. In addition, we may use deception to either deceive or assist others, such as lying to someone to avoid hurting their feelings or fearing their reaction if we tell them the truth. These actions require us to infer the unobservable mental states of other people's minds by observing their behaviour and assuming that others have a mind. Making this kind of assumption is relatively easy. In addition to attributing mental states to others, we are also capable of monitoring our own mental states. This ability is known as meta-cognition, and it allows us to attribute mental states to ourselves. For instance, when we encounter a complex task and feel confused, we can exercise meta-cognition by recognizing this state and taking steps to clarify our understanding, such as conducting further research or seeking assistance. Similarly, when assessing our confidence in a decision, we reflect on our knowledge or judgment in a given situation. Recognizing our emotions, such as happiness, sadness or frustration, and understanding how they may influence our thinking and behaviour is also an example of applying meta-cognition. The results of meta-cognition can be described and reported using natural language. The existence of minds in others is perceptible through communication and mutual understanding. However, the question of whether also non-human animals possess mental states remains uncertain, as they lack the ability to communicate with us through language. It

---

[1] In the cognitive sciences, ToM is often referred to as 'mindreading' or 'perspective taking.' It is also commonly known as 'folk psychology,' 'naive psychology,' or 'intuitive psychology.' The latter three terms indicate that this skill is used in everyday life by everyone, without the need for specific training in cognitive science. ToM is a widely recognized cognitive ability that is utilized spontaneously and unconsciously in our daily interactions with others. It is not a specific theory or model that we consciously apply.

is impossible to directly ask them if they have minds. It has to be notice that this uncertainty applies to all non-human animals, as well as prelinguistic infants.

The question of whether non-linguistic animals are capable of having a mind, and consequently, thinking and reasoning, has been debated in the history of philosophy since its origins. It continues to be a focal point in various scientific fields today. The cognitive turn[2] in the behavioural sciences has led to the recognition of cognitive processes in animals beyond human linguistic abilities[3]. This has been accompanied by interdisciplinary interest, particularly from fields such as cognitive ethology, developmental psychology, and cognitive archaeology, as well as ecology, neuroscience, and evolutionary biology. Recent scholarly consensus, informed by extensive research in various disciplines, suggests that many non-linguistic animal species exhibit behaviours and problem-solving abilities that require cognitive processes. For instance, some birds and primates have demonstrated problem-solving skills, such as using tools to obtain food, which indicates an understanding of the environment and objects within it (Auersperg et al., 2011; Seed & Byrne, 2010). Similarly, some octopus species exhibit behaviours that suggest planning and foresight, such as carrying coconut shells to use as shelter later (Godfrey-Smith, 2016). Additionally, studies on elephants and dolphins have demonstrated that these animals not only pass the mirror test for self-awareness[4], but also display empathy and social learning, indicating a high level of social cognition (Reiss & Marino, 2011; Plotnik et al., 2006; Marino, 2002). In his book

---

[2] The 'cognitive turn' in the behavioural sciences refers to a paradigm shift that occurred around the middle of the 20th century. Scientists began to recognize and study cognition as central to understanding animal behaviour. The cognitive turn introduced the idea that animals, including humans, possess internal mental processes that influence their behaviour. These processes can and should be studied scientifically. The emergence and expansion of cognitive psychology and cognitive science as fields of study have been influenced by behaviour research on animals. This has led to a greater focus on animals' cognitive abilities and an understanding of the mechanisms underlying their behaviour.

[3] For a comprehensive analysis of this subject, refer to Sorabji's (1995) works on historical philosophical perspectives, Bermúdez's (2003) insights into cognitive ethology, developmental psychology, and cognitive archaeology, and Vallortigara's (2000) contributions from ecology, neuroscience, and evolutionary biology.

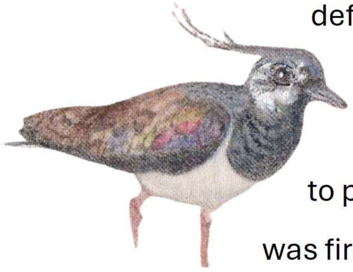
[4] The mirror test, also referred to as the mirror self-recognition test (MSR), is a behavioural measure created by psychologist Gordon Gallup Jr. in 1970 to evaluate self-awareness in animals. The test requires marking an animal with a visible mark in a location that is not visible to them without the use of a mirror. Once the mark is made, the animal is provided with access to a mirror. If an animal uses a mirror to investigate and possibly remove a mark, it indicates that the animal recognizes its reflection as an image of itself. This behaviour suggests self-awareness, as it implies that the animal understands that the reflection is not another animal but a representation of its own body. The mirror test is considered as evidence of self-recognition and has been successfully completed by only a few animal species, including humans, some great apes, dolphins, elephants, and possibly some birds. This indicates a high level of cognitive complexity.

'Thinking Without Words', Bermúdez outlines the key steps that contributed to the breakthrough in animal thought research.

Cognitive archaeologists have found evidence indicating that early hominid species engaged in thought-driven activities prior to the development of language. This challenges traditional theories about the timeline of cognitive evolution (Corballis & Lea, 1999; Mellars & Gibson, 1996; Mithen, 1990, 1996). This evidence suggests that early humans were capable of engaging in complex social coordination, such as adhering to seasonal hunting patterns, through a form of information transmission that went beyond the simplistic rules of primitive social structures. Additionally, it is hypothesized that the minds of our ancestors were equipped with specialized cognitive modules for interacting with specific elements of their environment, from tool-making to the establishment of social norms and practices. Mithen (1996) proposes that primitive rituals were not solely cultural artifacts, but rather derived from innate cognitive frameworks. Therefore, language likely acted as a catalyst that facilitated the integration of previously distinct cognitive modules, rather than enabling these complex activities.

Research in developmental psychology has extensively documented that prelinguistic infants possess an intuitive understanding of physical laws, even before they acquire language (Baillargeon, 1995; Spelke, 1990; Gopnik & Meltzoff, 1997). For example, studies have shown that around 5 months of age, infants exhibit signs of surprise when an object they expected to continue its trajectory disappears behind another object instead. This passage describes how infants demonstrate early conceptual understanding of object permanence. At around 6 months, infants display surprise when they see an object suspended in midair without any visible support, suggesting a developing comprehension of gravity. The observations suggest that infants engage in rudimentary thinking even without language skills. They form 'expectations' that, although not articulated through language, can be considered a fundamental aspect of cognitive processing. This challenges the idea that language is a prerequisite for thinking, illustrating instead that cognitive development begins with preverbal conceptualizations of the physical world.

Cognitive ethology has transformed our comprehension of animal behaviour by proposing that animals, like humans, are guided by their desires and beliefs about their surroundings. This shift in perspective moves away from the oversimplified view that animal behaviour is solely reactive to stimuli. Instead, it suggests that animals make decisions based on their desires, which are influenced by their beliefs, indicating a previously unrecognized level of cognitive complexity. The animal kingdom exhibits cognitive abilities that are widespread, as evidenced by behaviours that are interpreted as deliberate attempts to influence or deceive, not only within their own species but also in other species. Such behaviours have been observed across the phylogenetic spectrum. For instance, primates, including chimpanzees, have been documented to engage in sophisticated social interactions (Byrne, 1995). Primates live in complex social groups and engage in coordinated activities, communication, and both friendly and competitive interactions that vary with individuals and context. They form temporary alliances through mutual support, sharing food, and grooming each other. Their social behaviour is believed to have evolved primarily as a defense against predators, with additional benefits including improved resource identification, cooperative behaviour, and enhanced social learning. Similarly, plovers have been observed using a deceptive tactic to protect their eggs or young from predators. This anti-predatory strategy was first described by Ristau in 1991. They exhibit a distraction display by feigning an injury, such as a broken wing, to appear vulnerable and attract the attention of a potential predator. This lures the predator away from the nest. Once the predator has been led far enough away, the bird miraculously 'recovers' and flies away, leaving the confused predator behind. The titmouse emits false alarm calls, which can manipulate the behaviour of nearby birds to its advantage (Møller, 1988). These calls indicate the presence of a predator and vary in intensity depending on the perceived threat level. Titmice can produce these calls even without the presence of a predator, causing other birds to flee and potentially reducing competition for resources. However, this behaviour could pose risks. The constant issuance of false alarms could cause other birds to ignore its calls in the future, decreasing the effectiveness of genuine alarms. Finally, the behaviour of sentinels demonstrates cooperative vigilance in mixed avian communities. In these communities, birds designated as 'sentinel' assume the role of lookouts for predators while others forage (Munn, 1986). These sentinels perch in strategic



locations to monitor for danger, alerting the group to threats with specific calls that enable a rapid collective response. This system of shared vigilance promotes not only individual survival but also enhances overall group safety and success, reflecting the complexity of social structures and communication.

Thanks to the aforementioned studies, two significant changes have occurred in cognitive science. (i) Firstly, the concept of thought has been theoretically separated from language. While a strong correlation between thought and language was once widely accepted[5], current theories now recognize the possibility of non-linguistic thought. (ii) Researchers have developed experimental protocols that explore the nature and limits of non-linguistic cognition without relying on verbal responses. In recent decades, the focus of research has shifted from questioning the mere existence of mind in non-linguistic animals to a more detailed exploration of their cognitive abilities, including ToM. For over 45 years, scientists have debated whether non-linguistic animals can perceive the behaviours of their peers as guided by mental processes or simply as observable facts. While some researchers present substantial evidence to support the idea that some animals possess ToM, others refute this claim, citing Lloyd Morgan's canon[6] of simplicity as the basis for their skepticism.

---

[5] One of the most influential philosophical perspectives on the nature of thought is the idea that understanding thoughts is intrinsically linked to the analysis of the sentences used to express them. This perspective emerged from the works of Gottlob Frege, who is considered one of the founding fathers of analytic philosophy and the philosophy of language. Frege promoted the idea that thoughts are essentially abstract objects that can be expressed through language. Therefore, in order to fully comprehend a thought, it is necessary to dissect and analyze the logical structure of the sentences that constitute its expression. This conception establishes an inseparable link between thought and language, in which the meaning of a thought finds its fullest manifestation in its linguistic articulation. However, there are other perspectives that offer different nuances on the interaction between thought and language. Herbert Paul Grice proposes an approach that focuses on human communication and the pragmatic meaning of words within a conversational context. Unlike Frege's view, Grice emphasizes the importance of communicative intentions in the interpretation of meaning. This approach prioritizes the speaker's intended meaning over the literal meaning of the words used. Grice's theory is significant in the field of linguistics and has influenced the study of language and communication. Grice argues that understanding a statement involves not only the literal meaning of the words but also the inference of the interlocutor's hidden intentions. This allows for a richer and more multifaceted understanding of the communicated meaning. Similarly, Russell shares Frege's perspective that language is a vehicle through which thoughts can be expressed, but he emphasizes logic and mathematics. Russell aimed to resolve contradictions in logic and mathematics, which led to the creation of the 'theory of descriptions.' This theory explains how sentences containing definite descriptions can have meaning even when the object described does not exist.

[6] According to the canon of C. Lloyd Morgan, a British ethologist and psychologist, animal behaviour should not be interpreted as the product of complex mental processes or cognitive abilities if it can be explained by simpler processes or abilities. This principle is a restatement of Occam's razor, also known as the principle or law of parsimony. It suggests that simpler explanations should be preferred over more complex ones when analyzing natural phenomena from an ethological perspective. The importance of Morgan's canon is exemplified by the case of Clever Hans, a horse known for his supposed ability to answer simple mathematical questions by

Lurz (2011) points out that the ToM debate has reached an impasse, with neither side able to present conclusive evidence for their positions. Furthermore, the numerous empirical studies aimed at exploring mind-reading abilities in non-linguistic animals have not provided a definitive answer. This is because data from these studies can often be explained by simpler cognitive behaviours, leaving the central question unresolved. To move forward, researchers must address two critical issues: (i) a metaphysical inquiry into the nature of mind and the mechanisms underlying Theory of Mind (ToM); and (ii) an epistemological challenge regarding the criteria used to attribute ToM to non-linguistic animals. It is crucial to design experiments that would unambiguously demonstrate the presence of ToM in animals if successful.

While it is true that the cognitive sciences have freed the possibility of having a mind from the need to have language, the same cannot be said for ToM. Some argue that ToM in non-linguistic animals is tied to the possession of a public language[7]. It is important to note that this is a debated topic among scholars. Language is considered a necessary condition for the development of cognitive abilities such as ToM, meta-cognition, and logical reasoning. It is through language that we can be certain of the presence of these abilities. To attribute cognitive abilities in the absence of language, one alternative is to analyze the communicative systems of non-linguistic animals. This can provide support or evidence for species-specific features. Detractors of ToM in non-linguistic animals argue that different communicative

---

tapping his foot or shaking his head. Wilhelm von Osten, the owner of Clever Hans, claimed that the horse understood the concept of numbers and calculus. However, further investigations revealed that Clever Hans was not performing mathematical calculations, but rather responding to subtle involuntary cues from his owner or other observers. Clever Hans relied on nonverbal cues, such as changes in posture, muscle tension, or facial expressions, to determine when to stop tapping his foot or shake his head.

[7] The term 'public language' refers to a language that is used and understood by a community of speakers. Public languages are characterized by shared linguistic properties, such as vocabulary, grammar, and syntax, that enable communication and understanding among speakers. In contrast, a 'private language' is one that is known or used only by a single individual and is therefore not shared or understood by a larger community of speakers. Ludwig Wittgenstein proposed the concept of a private language in his book Philosophical Investigation. However, some historians of philosophy identify the work of Gottlob Frege and John Locke as precursors to this concept. Wittgenstein argues that a conception of language as private is necessarily incoherent, and therefore, such a language cannot exist. An example of a private language is the language of thought or mentalese. This language has characteristics similar to spoken language and plays the role of the actual structure of thought. According to this view, mental states are true propositions with a combinatorial syntax. This means that they have simple elements that combine to form complex elements, much like words combine to form sentences. Additionally, they have a compositional semantics, meaning that their meaning depends on the combination of the meanings of their constituent elements and how they are organized with each other. The thesis of the language of thought has a long-standing tradition dating back to late medieval philosophers. This idea was later revived by philosopher Jerry Fodor with the publication of his book, The Language of Thought (1975).

systems lack the necessary features to replace language as a necessary condition for the development or attribution of mindreading ability. But it is precisely this analysis that is the focus of critics of ToM in non-linguistic animals, who argue that different communicative systems do not have the necessary features[8] to replace language in its task as a necessary condition for the development or attribution of mindreading ability.

One problem with analyzing the cognitive differences between humans and other animals is the use of a unimodal approach that focuses solely on spoken vocalizations, rather than considering performed gestures, chemical signals, or different facial expressions in an integrated manner. It is important to recognize that all communication systems, including language, are multimodal. That is, they are systems that use different signals on different sensory channels simultaneously. For instance, primates use alarm calls to warn others of predators, and they also communicate through gestures like lip movements and teeth chattering. While these modes of communication have been studied separately, they are both important for understanding primate communication. However, primates often integrate visual and auditory communication, and fewer studies consider this feature. Ignoring this fact can lead to incorrect theories about the nature of the mind and hinder experimental protocols based on multimodal communication to test ToM.

A multimodal approach to studying the communicative systems of animals, both human and nonhuman, may be a potential solution to the stalemate in the ToM debate. This thesis focuses on addressing the epistemological problem (Chapters 3 and 4) while briefly touching on the metaphysical problem (Chapter 5). Indeed, I propose experiments to test the validity of my thesis that some behaviours related to multimodal communication require ToM. Thus, the relationship between multimodality studies (Chapter 2) and ToM studies (Chapter 1) is

---

[8] Examples of necessary features include recursiveness and compositionality. Recursiveness is a concept introduced by Noam Chomsky that refers to a language's ability to generate an infinite range of complex sentences through the iterative application of grammatical rules. An example of recursiveness is the construction of complex subordinate sentences. For instance, the sentence 'the cat that chased the mouse that ate the cheese that was in the pantry' uses recursion to include multiple subordinate sentences within one another. In contrast, Gottlob Frege developed the concept of compositionality, which states that the meaning of a complex linguistic expression is determined by the combination of the meanings of its individual parts and the syntactic structure that connects them. For instance, in the sentence 'the cat sleeps on the carpet,' the meaning of the sentence is derived from the combination of the meanings of the words 'cat,' 'sleeps,' 'on,' and 'carpet,' together with the syntactic structure that establishes the link between them. Non-linguistic animals may have similar features, but they are typically less developed than in humans or completely absent.

addressed for the first time. The following section provides a brief account of the importance, history, and gaps of ToM studies in non-linguistic animals. Section 1.1 discusses the implications of these studies in various fields of cognitive science. Sections 1.2 and 1.3 summarize the contemporary history of the debate surrounding human and non-human non-linguistic animals. Finally, section 1.4 aims to clarify existing gaps in ToM research by identifying areas that require further exploration to deepen our understanding of cognitive processes in non-linguistic entities.

1.2 Significance of Theory of Mind studies for cognitive science

Adopting a multimodal approach to investigate the communicative systems of non-linguistic animals could provide new insights into their ToM abilities. This topic has been the subject of significant debate. To fully understand the impact of this perspective, it is essential to consider how ToM studies intersect with and influence various disciplines in the cognitive sciences. These studies have a significant impact on cognitive ethology, developmental psychology, cognitive archaeology, and philosophy, shaping our understanding of cognition across species and developmental stages. Each field benefits from these insights, leading to a more integrated and comprehensive view of cognitive processes in human and nonhuman animals.

In the following section, we will discuss how ToM research initially gained traction through experiments on nonhuman animals after the decline of behaviourism[9]. Animal behaviour

---

[9] Behaviourism (or behavioural psychology) is an approach to psychology that focuses on the study of an individual's explicit behaviour as the only scientifically studiable unit of analysis. It was developed by psychologist John Watson in the early twentieth century and represented a significant break with the earlier approach of classical psychology, which focused primarily on the study of consciousness, sensations and perceptions. According to behaviourism, psychology should exclude internal events and mental states, viewing them as a "black box" that cannot be accessed directly. Instead, it focuses on observable behaviour, which manifests itself as an individual's reactions in certain situations, and the external stimuli that determine their responses. Behaviourism emphasizes the role of the environment in shaping an individual's behaviour through conditioning, where environmental stimuli are associated with specific responses. Reinforcement, whether positive or negative, can increase or decrease the likelihood of a behaviour recurring. One famous example of behaviourism in action is Watson's experiment with 'Little Albert.' In this experiment, a 9-month-old child was conditioned to associate a neutral stimulus (a white mouse) with an unpleasant stimulus (a loud noise). As a result of this conditioning, the child began to cry and exhibit signs of terror whenever he saw the white mouse or similar objects. Behaviourism has had a significant impact on psychology by providing a scientific basis for the study of human behaviour. However, it has also received criticism for its lack of consideration for the internal aspects of the human mind and emotions. Additionally, some have objected to the use of experimental

scholars needed a valid methodology to explain behaviours that could not be accounted for as an invariant response to specific stimuli. A behaviour is defined as an invariant response to specific stimuli when it is an innate release mechanism. This concept was developed by Konrad Lorenz and Nikolaas Tinbergen, along with the concepts of fixed pattern of action and sign stimuli (Ronacher 2019). This mechanism is characterized by being triggered by specific stimuli, always occurring in the same way, being present in all members of the same species, not depending on an individual's personal history, not being able to be interrupted or modified once triggered, and having only one specific function (Lorenz 1935, Lea 1984)[10]. This is exemplified by the wild goose's egg retrieval behaviour. Like many ground-nesting birds, geese use their beaks to return eggs that have moved from the nest back to their original position. The process involves fixing their gaze on the egg, stretching their neck over it, and rolling it back to the nest using the lower part of their beak. The goose will continue this behaviour even if the egg is removed during the rolling process. When a behaviour, such as that of an egg-rolling goose, cannot be explained by innate release mechanisms or other biological processes, it requires an explanation from the field of psychology. This approach involves attributing the behaviour to cognitive processes, suggesting that the animal's mental state may be the underlying cause. For instance, numerous animals, particularly juveniles, exhibit playful behaviour. Puppies chase each other or play with toys, and some primates engage in similar activities. Additionally, some species employ tools to accomplish specific objectives, such as obtaining food. These behaviours suggest motivations for exploration, social interaction, and enjoyment, or demonstrate an awareness of how certain objects relate to achieving specific goals. Such complexities exceed what can be attributed to mere physical or automatic responses to environmental stimuli. Several methodologies have been developed to attribute mind in non-linguistic animals (Bermùdez 2003; Griffin, 1992; Shettleworth,

---

techniques that may be considered ethically questionable, such as the Little Albert experiment. Despite receiving criticism, behaviourism played a crucial role in developing a scientific perspective in psychology and paved the way for new approaches to studying human behaviour. Other psychologists, such as Clark L. Hull and Edward C. Tolman, also contributed to further developments in behaviourism, expanding its field of research and application to different areas of psychology.

[10] It should be noted that the term 'innate' is controversial because it can refer to multiple characteristics, such as behaviour that is present at birth, unlearned, developed before being used, unchanging once developed, adapted in the course of evolution, served by a specific module in the brain, or attributable to some genetic trait. On the other hand, some studies suggest that behaviours we commonly consider innate may actually develop from species-specific ecological experiences and contexts (ontogenetic niche), as well as inheritance of genetic traits (Ewert 2013, Blumberg 2017).

1998). However, the search for a comprehensive and reliable set of methods to ascertain the extent to which these animals possess ToM is ongoing and remains a critical gap in the field. Different approaches to the study of ToM contribute to meeting that need. The study of ToM enhances ethologists' comprehension of social behaviours in non-linguistic animals, encompassing cooperation, competition, communication, leadership, and altruism. Additionally, it offers insights into how these social dynamics adapt to environmental changes, such as those resulting from urbanization and climate change. Findings in cognitive ethology can help identify the presence and extent of cognitive abilities in different animal species, delineating common characteristics and developmental stages. Additionally, this field enhances our understanding of human cognition by offering insights and theoretical models that inform developmental psychology (Ma et al., 2023). These models are important for studying the development of cognitive processes in prelinguistic infants and in individuals with conditions such as autism spectrum disorder (ASD) or schizophrenia. Additionally, cognitive ethology aids in reconstructing the evolutionary history of cognitive mechanisms, providing insight into how these processes may have originated and evolved.

ToM is a cognitive ability that typically develops in early childhood, reaching maturity around the age of four. However, it may be impaired or not fully developed in individuals with certain conditions, including depression and anxiety[11] (Wolkenstein et al., 2011; Lee et al., 2005; Wang et al., 2008; Washburn et al., 2016). Ethology has renewed interest in studying ToM, but developmental psychology has paved the way for experimental methods to track ToM progression in humans and other species. (i) For instance, Joseph Perner and Heinz Wimmer (1983) developed the false belief test, also known as the Sally-Anne test, as a crucial tool for evaluating ToM. In the classic Sally-Anne false belief test, a scenario is presented to a child using two dolls as props. The dolls in the story are Sally and Anne. Sally has a basket, while Anne has a box. The story goes that Sally puts a marble in her basket and leaves the room. While she is gone, Anne moves the marble from the basket to her box. The question asked of the child is where Sally will look for the marble when she returns. This assessment evaluates a child’s ability to comprehend that Sally, based on her beliefs, will continue to think that the marble is in the basket, despite its actual location. A correct response indicates an

---

[11] Section 1.3 will provide a more detailed explanation of these aspects and their implications for ToM.

understanding of Sally's false belief, which is a key aspect of ToM development. An incorrect response suggests that the child may not have yet grasped the concept that others may hold beliefs that differ from reality. (ii) The experience projection test, often associated with the work of Andy Meltzoff (2008), assesses how children use their own experiences to infer the experiences of others. In one experiment, 12-month-old children wore a blindfold themselves, experiencing firsthand its ability to obstruct vision. These children, now familiar with the effect of the blindfold, demonstrated a nuanced understanding that a person wearing a blindfold cannot see. The study showed that the experimental group of children had a reduced tendency to follow the gaze of a blindfolded adult turning toward an object, unlike the control group. The control group consisted of children who had interacted with the blindfold (by seeing or touching it) without experiencing vision obstruction, and those who had worn a modified blindfold equipped with windows that did not obstruct vision. (iii) Another common approach to assess ToM in children is the deceptive content test, also known as the Smarties test. In this scenario, a child is presented with a familiar container, such as a box of Smarties, that unexpectedly contains a different object, such as pencils. The child is then asked to predict what another person who has not seen the contents of the box would expect to find inside. If the child understands that the other person would be fooled into expecting Smarties instead of pencils, it indicates that they are capable of comprehending false beliefs. The false belief test, experience projection test, and deceptive content test are important methodologies for assessing ToM. However, they represent only a subset of the various experimental protocols available to explore this cognitive ability. There is a wide range of methods designed to empirically investigate ToM in different age groups and contexts that go beyond these two approaches, and their accuracy and applicability may vary (Chapter 4). Ongoing research on ToM abilities is essential for developing new experimental methodologies and enriching the pool of diagnostic tests for ToM. These advances are crucial for further investigating the typical functioning and evolution of ToM in both humans and some nonhuman animals. They also aid in exploring the neurobiological basis of cognitive disorders previously discussed (Hoerold et al. 2008; Robinson, Hertzog, & Dunlosky 2006; Shimamura & Metcalfe 1994). The knowledge gained from this research is crucial for developing therapeutic strategies for these disorders.

Developmental psychology, cognitive ethology, cognitive archaeology, and evolutionary biology have collaborated to study the origins and evolution of ToM and other related social-cognitive skills (Emery & Clayton, 2001; Santos et al., 2007; Penn et al., 2008). Cognitive archaeology enriches the investigation of intuitive understanding of physics or biology[12], exploring the hypothesis that some domain-specific intelligences[13] may be uniquely human traits. Cognitive archaeology utilizes observations of nonhuman primate social interactions to hypothesize about the social behaviours of our ancestors. These insights have led to the development of new methods of inquiry in developmental psychology, such as the habituation/dishabituation paradigm[14]. This paradigm has been proven invaluable for understanding cognitive processes in both human infants and nonhuman animals (Hauser, 1998; Munakata et al., 2001). In the context of infant cognition, researchers often use this approach to study the perceptual and cognitive abilities of infants and young children. The ability of individuals to habituate and dishabituate to visual, auditory, or tactile stimuli provides a valuable means of understanding learning and discrimination processes during early development. Similarly, the habituation/dishabituation paradigm is being used in animal mind studies to explore the cognitive abilities of other species, such as nonhuman primates. Observing an animal's ability to distinguish between familiar and novel stimuli can provide valuable insights into their perception and understanding of the world around them. In

---

[12] Similar to naive psychology, in the field of perception and understanding of physical phenomena, there is a concept known as 'folk physics' or 'naive physics.' These terms refer to people's intuition and innate conceptions about the properties and physical principles of the world around them, such as gravity, motion, or solidity, without specific training in physics. For instance, the expectation that a dropped object will fall downward. Similarly, in the context of understanding biological phenomena, the terms 'folk biology' or 'naive biology' are used to refer to the innate beliefs and intuitions that people hold about living organisms, their functioning, and interactions. For instance, the concept that animals require food to grow or that plants require sunlight are examples of concepts belonging to naive biology. These ideas often arise in our daily lives and influence how we interact with the natural world. However, they may not always align with formal scientific knowledge.

[13] "Domain-specific intelligences" refer to cognitive abilities that are specialized to handle specific types of information or problems within certain domains. These intelligences are believed to have evolved in response to specific environmental or social challenges that our ancestors encountered. As a result, dedicated cognitive mechanisms have developed to efficiently handle these specific tasks. For instance, a domain-specific intelligence may be related to the ability to recognize and remember faces (social domain) or the ability to perceive and reason about the trajectory of moving objects (physical domain). Domain-specific intelligences are distinct from general or domain-general intelligences, which are cognitive abilities applicable to a wide range of problems and situations. In this chapter, we refer to Mithen's theories and advances in cognitive archaeology in relation to these types of intelligences.

[14] This methodology is based on the ability of humans and other animals to recognize and respond to new information in their surroundings. Initially, individuals are exposed to a repeated stimulus, a process known as habituation, which leads to a reduction in the initial response to that stimulus. However, through dishabituation, a slightly different or completely new stimulus is introduced. If the individual is able to detect the difference between the habituated stimulus and the new stimulus, a renewed and more vigorous response occurs. This change in response suggests that the individual has recognized the novelty of the stimulus.

evolutionary biology, studies on cognitive abilities are crucial to understanding the development and impact of these abilities on social adaptation and cooperation in the animal kingdom. They provide insight into the adaptive advantages linked to different cognitive functions, improving our comprehension of the factors that contribute to evolutionary success. Studying ToM in nonhuman primates provides insight into the evolution of this ability within their lineage, as well as that of hominids and potentially even more distant ancestors. This research can also help identify the specific evolutionary pressures that shaped the capacity for ToM in different evolutionary lines. Comparative cognitive studies, which are enhanced by neuroimaging techniques such as functional magnetic resonance imaging (fMRI), allow for the real-time mapping of brain activity. Additionally, electroencephalography (EEG), which records the electrical activity of the brain, facilitates the analysis and comparison of features and functions of the nervous system in different animal species. These studies can reveal the brain regions involved in ToM, such as the medial prefrontal cortex or the mirror neuron region and enable us to test whether similar neuronal activity is present in non-human animals. Simultaneously, research on mindreading in nonhuman animals provides evidence of the presence of this cognitive ability in different species. This allows cognitive science to identify nonhomologous but human-like brain regions that support ToM activity, which is crucial for understanding the development and evolution of cognitive abilities.

In philosophy, the study of ToM and cognitive abilities has two main implications. (i) Firstly, it influences discussions on ethics. (ii) Secondly, it informs debates on the nature of the mind. The recognition of cognitive abilities similar to those of humans in other animals raises important ethical questions. This is particularly relevant to animal welfare, as some ethicists link the moral relevance of animals to their level of cognitive sophistication. The discovery of mindreading activity in non-linguistic animals necessitates greater attention to their needs to ensure that their living conditions and treatment respect their cognitive states. This principle applies to both ex-situ conservation efforts and experimental practices, which are subject to varying regulatory practices across different countries[15]. Some of these regulations may

---

[15] For instance, the European Union's animal experimentation principle, the '3Rs' approach, emphasizes 'Reduction' (using fewer animals), 'Refinement' (improving experimental techniques to minimize discomfort or suffering), and 'Replacing' (using alternatives to animal testing). These guidelines have been incorporated into the EU's legislation on the protection of animals used for scientific purposes. If future research confirms the

consider ToM as a valid indicator for imposing restrictions or banning animal experimentation involving animals with this capacity. When considering the nature of mindreading, ToM studies have implications for both (i) the relationship between language and thought, as well as (ii) the role that mindreading abilities play in the cognitive functions of animals, both human and nonhuman. On one hand, some argue that ToM develops independently of language and its evidence is only possible at a certain level of language development (Fodor 1992; Chandler, Fritz & Hala, 1989), or that language development depends on ToM (Grice 1975). On the other hand, some theorists argue that language is crucial for the development of ToM (Frye, Zelazo & Palfai, 1995; Gopnik & Wellman, 1994; Perner, 2000; Dunn, Brown, Slomkowski, Tesla & Youngblade, 1991). They argue that it is through various communicative contexts that the mind is formed, leading to the acquisition of concepts[16] such as beliefs, desires, and intentions (Bartsch e Wellman, 1995; Olson, 1988; Peterson e Siegal, 2000; Bretherton e Beeghly, 1982). Robert Lurz (2015) groups existing philosophical theories on the role of ToM ability in the development of cognitive functions into three categories. (i) The first category examines whether ToM enables the acquisition of the concept of objectivity relative to a state of affairs (Wittgenstein 1953; Davidson, 1980, 2001). (ii) Others have questioned whether ToM is responsible for the ability to think about specific objects and their properties (Strawson 1959; Welker 1988; Proust 2009; Burge 2009, 2010). (iii) Similarly, some have asked whether ToM is a necessary requirement for metacognition (Carruthers 2000, 2009; Tomasello 2023). The different positions on these questions stem from the historical tendency in

---

presence of mindreading capabilities in certain animal species, experimental policies and practices may need to be revised to account for these new findings. An understanding of advanced cognitive abilities could impact the application of the 'Refinement' principle, necessitating greater attention to experimental conditions and animal welfare. The presence of mindreading capabilities could also lead to increased application of the 'Reduction' principle, limiting the use of such species in experiments or imposing further restrictions based on their specific cognitive needs.

[16] Concepts are mental representations that group objects, events, or ideas into categories based on common characteristics. They are fundamental constructs in the way animals organize and understand the world around them. Each concept represents a generalization of similar objects or ideas, enabling individuals to process information efficiently and draw inferences based on similarities between items within the same category. Cognitive constructs are crucial for thinking, learning, and communication. When acquiring new concepts, individuals associate key features of objects or ideas into a category and develop a mental model that represents the essence of that category. For instance, the concept of 'animal' may include various creatures, such as dogs, cats, and birds, that share common properties, such as the ability to move and the need for food. The capacity to generate, employ, and distinguish concepts is crucial for comprehending language, resolving problems, making decisions, and acquiring knowledge. Concepts enable us to classify intricate experiences and information into significant categories, streamlining cognitive processing. Analyzing concepts enables us to detect connections, identify patterns, and predict outcomes based on prior experiences.

philosophy and science to identify a unique ability in humans that distinguishes them from other animals. Therefore, claiming that ToM depends on language, a skill typically associated with humans, reinforces the narrative of human superiority. It is important to maintain objectivity and avoid subjective evaluations. Research on the capacity for mindreading in other animals contributes to shaping our understanding of the nature of mind and potentially changing the anthropocentric narrative in favor of an evolutionary ecological view of species development. If non-linguistic animals were shown to possess ToM, it would demonstrate that this cognitive capacity, along with others that rely on ToM, is not language-dependent. This would remove it from the domain of human species-specific cognitive faculties and contribute to the development of new theoretical models of the mind.

1.3 The History of Theory of Mind Studies in Non-Human Animals

The challenge of attributing a ToM to non-linguistic animals has been a significant issue for both philosophy and science for centuries. Historically, figures such as Descartes and Hume have had a profound impact on the formation of Western philosophical and scientific perspectives on animal minds, touching on both metaphysical and epistemological issues. Descartes (1637) is often considered the philosopher who denies any form of cognition, such as reason, thought, or consciousness, to animals. This is due to his well-known ontological dualism, which reduces animals to mere biological machines. In contrast, Hume argues for the existence of animal cognition in his 'Treatise on Human Nature' (1739-40). While Hume acknowledges a distinction between human and animal minds, he emphasizes that the differences are only a matter of degree, not of kind. The author presents a metaphysical view of animal minds, suggesting that animal cognition is fundamentally similar to human cognition. Additionally, the author provides an epistemic justification for understanding other minds.

To establish a clear demarcation within the contemporary history of these studies, the 1978 landmark paper by psychologists David Premack and Guy Woodruff serves as a benchmark.

The paper, entitled 'Does the Chimpanzee have a Theory of Mind?'[17], tested whether chimpanzees are able to infer mental states through innovative experiments. This study is significant due to its exploration of the ability of chimpanzees to understand mental states. In a significant experiment inspired by Kohler's earlier work (1917), chimpanzee Sarah was shown videos of a human actor dealing with various problems. After each video, Sarah was presented with two photographs and asked to choose between them. One photograph showed the actor performing an action that would solve the problem depicted in the video, while the other showed an unrelated action. For instance, if the video demonstrated the challenge of reaching for a banana, one photo could display the actor using a stick to retrieve it (the correct solution), while the other could show the actor walking away. Sarah consistently selected the correct photo, which indicated the solution to the problem, in 21 out of 24 trials. This result was significant enough to suggest that she comprehended the actor's intentions and could distinguish which action would fulfill them. Several researchers (Savage-Rumbaugh et al. 1978; Heyes 1998; Povinelli, 1999) challenged this hypothesis over time. Upon closer analysis of the entire survey method, they concluded that the results proposed by Premack and Woodroof were experimental artifacts. Specifically, they contested that Sarah's ability to choose the most suitable picture for the video she had observed did not depend on her ToM ability, but rather on two factors: (i) the intrinsic characteristics of the presented pictures and (ii) her previous experience with other experimental protocols. In the first case, Sarah could have used visual element matching strategies to make her choices. For example, she could have chosen a picture where the actor used a stick to solve a problem, simply because the same object was present in the last frame of the video. In the second case, however, the participant may have relied on previously learned associations. For instance, they may have chosen the photo with a burning paper roll because they had learned to associate the orange flame with the operation of the radiator prior to the experiment. These claims were quickly challenged by further experiments that produced negative results regarding chimpanzees' understanding of both visual perception (Povinelli & Eddy 1996; Povinelli et al. 1994) and false belief perception (Call & Tomasello 1999). As a result, it was

---

[17] The term Theory of Mind was coined by them to refer to the ability to attribute mental states to oneself and others through a system of inferences, in order to explain and predict behaviour. The distinction between ToM and metacognition, which is the ability to monitor one's own mental states, occurred later.

concluded that nonhuman primates were unable to understand the mental states of others, unless clearly marked as subjective evaluations.

Experiments on visual perception aimed to determine if chimpanzees could comprehend and react to the caregiver's gaze direction, especially when used as a cue to request food. Caregivers were either positioned to visually interact with the chimpanzees or not. At first, the chimpanzees did not exhibit a significant preference for interacting with attendants who had a direct line of sight with them. Over time, the chimpanzees began to show a slight preference for seeking assistance from caregivers who were visibly attentive. To validate the experimental methodology, Povinelli and Eddy (1996) compared the chimpanzees' behavioural responses with those observed in preschoolers. They focused on the chimpanzees' ability to interpret and respond to visual attention cues from others. The researchers concluded that the chimpanzees' behaviour was related to learning rather than an understanding of the concept of sight. A nonverbal Sally-Anne test was conducted to investigate false beliefs in chimpanzees. During the experiment, one adult, referred to as the 'hider,' placed a reward in one of two containers. A second adult, the 'communicator,' observed the action and then attempted to assist the participants, which included children aged 4 to 5, chimpanzees, and orangutans, by placing a marker on the container in which they believed the prize had been hidden. At this stage, the participants demonstrated their understanding of the game by correctly choosing the container. In the second phase, the hider moved the prize to the other container while the communicator was away. Upon returning, the communicator was unaware of the exchange and placed the marker on the original, now empty container. To succeed in the game, participants had to correctly guess that the prize had been moved to the opposite container, despite the communicator's marker. However, none of the nonhuman primates were successful in the second phase of the experiment.

Prior to Premack and Woodruff's study, several studies had already been conducted on the evolution of cognition (Chance & Mead 1953; Jolly 1966; Humphrey 1976). These studies led to the development of the social or Machiavellian intelligence hypothesis. This hypothesis proposes that cognitive abilities, such as recognition of individuals, memory of previous social interactions, and theory of mind, evolve due to selective pressures exerted by collaboration and competition in social contexts. According to Brothers (1990), the ability to predict, manipulate, and learn from others develops as social systems become more complex. This hypothesis explains the association between high cognitive abilities and greater brain size in primates, as evidenced by the larger neocortex area. The hypothesis that remains the most predictive of the volume of specific areas of the central nervous system, not only for primates but for all mammals, is still in effect[18]. Inspired by the work of Premack and Woodruff (1978), as well as by the broader field of cognitive evolution, experiments began to focus on ToM related to social interaction in nonhuman primates. Specifically, researchers were interested in intentional deception by one individual towards another conspecific (Byrne and Whiten 1988; Byrne 1995; Whiten and Byrne 1997) and the use of alarm signals to warn members of the same social group of impending danger (Seyfarth, Cheney and Marler 1980). This study was motivated by earlier anecdotal references from the late 19th and early 20th centuries (Gardner 1892, 1896, 1900; Yerkes & Learned 1925).

Lurz (2011) identifies two studies as milestones in the history of ToM in non-linguistic animals due to their significant impact. The first study, conducted by Emil Menzel (1974), examined the spatial memory of captive chimpanzees. The second study, conducted by Dorothy Cheney and Robert Seyfarth (1990), investigated alarm signals in wild vervet monkeys. Menzel conducted an experiment in which food was hidden within a 4km^2 area. An adult chimpanzee was shown the location of the food and then returned to its group. The experiment tested whether the subject could find the food again. During the experiment, an adult female named Belle was able to consistently locate the hidden food by walking a straight line. Belle was followed by the alpha male of the group, Rock[19]. In a previous instance,

---

[18] Alongside the well-established 'social or Machiavellian intelligence hypothesis,' another strand of thought known as the 'ecological intelligence hypothesis' is emerging. The former highlights the complexities of social interactions as a driver of cognitive evolution, based on research and observations of complex primate social interactions. The latter emphasizes the importance of the challenges posed by the natural environment. This hypothesis highlights the potential importance of dietary characteristics, spatiotemporal distribution of foods, and the need to adapt to a changing environment in the development of primate cognitive abilities. While social complexity has traditionally been viewed as the primary driver of primate cognitive evolution, recent research suggests that ecology may be equally influential. Studies have shown that dietary niche can predict many of the same neurobiological characteristics as social complexity. Additionally, ecology has been demonstrated to influence cognition and behaviour in other species, such as birds.

[19] Chimpanzee communities have a clear hierarchy that determines behaviour and interactions among individuals. This hierarchy establishes who has control and authority within the group. Typically, there are dominant chimpanzees, often referred to as 'alphas,' who have the main power and hold control over resources such as food, territory, and mating opportunities. These alpha individuals often exhibit dominant behaviours and may use force or threat to maintain their position. On the other hand, chimpanzees can be classified as dominant or subordinate. Subordinate chimpanzees are generally more passive and exhibit submissive behaviours to avoid conflict with dominant individuals. This hierarchical dynamic is complex and influenced by

Rock had understood Belle's pattern and consumed the food without leaving any for her. During the experiment, Belle exhibited a new behaviour that had not been observed before. Instead of going towards the location where the food was hidden, she went in the opposite direction, with Rock following her. During the experiment, Rock was searching for food in the wrong place while Belle went back to the correct spot to consume her reward. This behaviour was interpreted by Menzel as an attempt to induce a false belief in a conspecific for personal gain. Following this experiment, others were performed (Cossui-Korbel 1994; Hirata & Matsuzawa 2001). Spontaneous observations of intentional deception were also collected in several primates, particularly in great apes (Byrne 1995), all reporting results similar to those obtained with Belle (Bryne & Whiten 1990).

Seyfarth and Cheney (1990) published their findings on alarm signals in wild vervet monkeys, building on their previous research (Seyfarth, Cheney and Marler 1980). The study revealed that these signals have semantic content and propositional nature, going beyond simple emotional or instinctive reactions. Guenons are known to produce vocalizations specific to the type of predator they encounter, such as snakes or eagles. They respond flexibly and appropriately to different types of alarms, climbing trees when they spot a snake or looking at the sky when they see an eagle. Seyfarth and Cheney recorded individual alarm signals emitted by specific individuals to demonstrate the semantic content of their vocalizations. Once the subject whose sounds were recorded departed from the group, the guenons began to play back the different sounds in the absence of danger. For instance, they played the alarm call for an eagle even when there was no eagle around, and recorded the responses from the other guenons. This experiment highlighted that the guenons stopped responding quickly to recorded alarm calls once they realized that these were unreliable. Not only did they stop following the single alarm signal, but they also stopped responding to the entire range of calls, even if they were directed towards different dangers, such as a snake, for example, as long as they came from the same individual. However, this did not occur if the signal came from individuals other than the one now considered

---

various factors such as age, sex, relationships, and alliances. However, it is crucial to maintaining order and structure within the community.

untrustworthy. The guenons demonstrated that their responses were not solely based on the characteristics of the signal, but rather on the meaning of the signal as conditioned by the individual emitting it. Seyfarth and Cheney's interpretation of this phenomenon was conservative, as they did not claim the presence of any form of intentionality. However, research results showed clear evidence to the contrary. These findings prompted several researchers to investigate the evolution of language, social organization, and communication systems of primates (Dunbar 1996). Sue Savage-Rumbaugh's study of Kanzi (1996) is iconic in this field. Kanzi, a young bonobo, was educated in communication with a strong emphasis on affectivity and interspecies social interaction with humans. Kanzi learned to understand English words and whole sentences, while developing a psychologically complex affective network.

With the advent of neuroimaging and advances in technology, researchers have gained new tools to explore the neurological basis of Theory of Mind (ToM). In the 1990s, a research group led by neuroscientist Giacomo Rizzolatti of the University of Parma made a revolutionary discovery of mirror neurons in animals, particularly in macaque monkeys (Gallese et al., 1996; Rizzolatti et al., 1996). This discovery has opened new avenues for the study of Theory of Mind not only in animals but also laid the groundwork for similar research in humans. Mirror neurons are a type of neuron that is activated both when an individual performs a specific action, such as grasping an object, and when observing someone else perform the same action. Mirror neurons are named as such because they reflect the actions performed by others, enabling observers to comprehend and imitate the actions. In essence, mirror neurons facilitate our understanding of the intentions and actions of others through an internal simulation of the action in our brain. Simulation theory suggests that mindreading involves using one's own mind to imagine and reconstruct another person's mental perspective. This is similar to completing an incomplete picture by filling in the missing details with our inferences and prior knowledge. Recently, researchers tested this theory on chimpanzees (Lurz et al. 2022) to determine their ability to comprehend and anticipate the actions of other individuals searching for a previously concealed object. During the experiment, the chimpanzees were subjected to two different conditions: (i) one in which another individual knew the location of the object relative to its original position (true-belief condition), and (ii) one in which the other individual did not know where the object had been moved (false-belief condition). The chimpanzees were then required to search for the object. The collected results indicate that chimpanzees rely more on other individuals' beliefs than their own when searching for an object, supporting the simulation theory as the underlying mechanism of mindreading. The utilization of neuroimaging and theories regarding the mechanisms that facilitate ToM do aid in addressing one of the primary inquiries of ethology, as proposed by Tinbergen (1963). Tinbergen argues for the necessity of distinguishing between proximate and ultimate causes in exploring animal behaviour. Proximate causes refer to the underlying mechanisms and ontogenetic development, while ultimate causes refer to the functional level and evolutionary history of the behaviour.

Four philosophers, Donald Davidson (1975, 1982, 1997), Ruth Millikan (2006), Fred Dretske (2006), and Jose Luis Bermudez (2003), have made significant contributions to our understanding of the nature of the mind and the functioning of Theory of Mind (ToM). They have steered the debate on animal mind reading in a specific direction by supporting the thesis known as the 'linguistic master argument,' as defined by Glock (2018). This argument states that thinking ability is dependent on the possession of concepts, which are based on language. Therefore, nonhuman animals, who do not possess language, would be incapable of thinking. This implies that nonhuman animals would also lack ToM, as it involves an inferential thought process. The same is true for other non-linguistic animals, such as infants. This position is the least optimistic regarding the cognitive abilities of non-linguistic animals. It was originally formulated by Davidson, with contributions from Stich (1979) and Dummett (2010). Dretske (2006) and Millikan (2006) argue that animals can have concepts such as beliefs or desires, but only perceptual ones. They still relegate the possibility of logical-formal and meta-representational reasoning to articulated language. Bermudez aims to reduce the disparity between human and animal cognition. While concurring with Millikan and Dretske that formal-logical reasoning is impossible without language, Bermudez proposes the existence of a primitive or proto-logical form in non-linguistic animals, drawing on the concept of proto-thinking postulated by Dummett (1993). Formal logic is based on well-defined rules and symbols, such as propositions, logical concepts, and truth values. Non-linguistic animals may exhibit a form of reasoning similar to formal logic, but based on visual, spatial, or sensory elements rather than propositional-type linguistic structure, logical concepts, and truth values. Using the works of these four philosophers as cardinal points, several rebuttals have been presented by other philosophers seeking to counter the arguments mentioned above. Some of these philosophers include Allen & Bekoff (1997), Amstrong (1973), Tye (1997), Glock (2000), and Carruthers (2008). They argue that it is possible to attribute mental states to animals, even if they are not aware of them. Carruthers (2009, 2013) and Hurley (2003) attempt to extend Bermùdez's argument by imagining a logic-rational capacity in non-linguistic animals. Meanwhile, Andrews (2015), Lurz (2007), Rescorla (2009), and Vigo and Allen (2009) argue for the possibility of non-linguistic animals engaging in inferences through different cognitive strategies.

In 2008 and 2022, Tomasello published two papers reviewing the progress made since the release of Premack and Woodruff's seminal paper (1978). The findings during this period demonstrate that great apes do not simply read and react to the behaviour of others, as previously thought. The reviewed evidence indicates that great apes understand their own and others' goals, intentions, perceptions, and knowledge. Moreover, individuals comprehend how various psychological conditions interact to generate intentional actions. They understand others based on a relatively coherent psychology of perception and goals, in which the other person acts in a particular way because of their perception of the world and their goals for how they want the world to be. Although the evidence supporting this claim is less consistent, it is possible that other nonhuman primate and bird species may also possess a similar understanding. However, they would not be able to understand false beliefs, which is a unique ability of the human species. In reviewing these studies then Tomasello hypothesizes a correlation between ToM and meta-cognition by postulating five hypotheses to further the understanding of social cognition in great apes and other primates (i) Among primate species, and perhaps also among non-primate species, there should be a correlation between self-regulatory abilities and social cognitive abilities (ii) Specifically, species that exhibit behavioural inhibition skills should be able to predict and control the behaviour of others (iii) species that demonstrate metacognition skills, such as information seeking in situations of uncertainty, should be able to attribute mental states to others, although this ability might be limited to species that also exhibit social learning skills and/or behavioural imitation (iv) Species that possess self-regulatory or metacognition skills related to social cognition, at any level, should engage in more intense food competition than related species (v) Although more speculative, the hypothesis suggests that the two skill sets are interdependent during cognitive processing, such that metacognitive monitoring, e.g., overloading it with distractions or tasks, may adversely affect social cognitive skills at that time.

1.4 A Brief Overview of Theory of Mind in Humans

Parallel to studies conducted with nonhuman animals, two conferences were held in Toronto and Oxford in 1986, organized by psychologists and philosophers Janet Astington, Alison Gopnik, and Paul Harris, to investigate ToM in humans. The studies began by involving children, then infants in prelinguistic age, and later integrated adult and elderly subjects. The first two aspects studied were (i) the ability to understand how a subject's beliefs - both true and false (Wellman & Bartsch, 1988) - and desires work together to produce intentional behaviours (Wimmer & Perner,1983). Imagine a child, Marco, who believes that his teddy bear is hidden under his bed. Marco wishes to play with his teddy bear. These two pieces of information - his belief and his desire - work together to influence his behaviour. Marco will search under the bed for his teddy bear because he believes it is hidden there. (ii) The capacity to distinguish between a mental entity and its physical counterpart (Wellman, 1990). For instance, an apple has objective properties such as color, shape, taste, and texture. However, perceptions of an apple's taste and texture, which constitute subjective mental entities, can significantly vary among individuals. For example, one individual may describe the apple as palatable and firm, indicating a positive sensory experience. Conversely, another individual might characterize the same apple as excessively tart or not sufficiently matured, reflecting a less favorable sensory assessment. The apple has objective physical properties, but our perceptions and thoughts about it are subjective. Children between the ages of 4 and 6 have shown competence in these two cognitive abilities, according to early studies.

Around the same time as the Toronto and Oxford conferences in 1986, researchers began to explore how ToM fits within an individual's learning, neural, and developmental processes. One hypothesis was that ToM depends on the organism's level of development. If this is true, then the development and progression of ToM abilities in humans should follow a universally applicable trajectory, unaffected by individual traits or cultural variations. To investigate these mechanisms, studies began to include not only typical subjects but also atypical subjects, such as individuals with autism spectrum disorders (ASD) or deafness. Additionally, more attention was paid not only to subjects' individual experiences but also to the cultural context in which they were embedded. As a result, it was discovered that the timing of ToM development varied. Children diagnosed with ASD typically develop an understanding of the concept of false belief at a later stage than children without this disorder (Baron-Cohen et al, 1985). This suggests a generalized delay in neurodevelopment, as other cognitive abilities also follow different developmental time frames in autism. However, additional research on deaf children revealed comparable outcomes despite the absence of the same neurological deficits found in children with ASD (Peterson & Siegal, 1995). Deaf children born to hearing parents were found to have delays in ToM development that were comparable to those of children with autism spectrum disorder. It took them about 12 years to reach the same stage of development that typical children reach in about one-third of the time (Peterson et al., 2005). To test whether cultural differences impact ToM development, comparative studies have been conducted between Western, Middle Eastern, and Eastern children (Wellmann et al., 2006; Shahaeian et al., 2011). The studies reveal that Iranian, Turkish, and Chinese children develop ToM in comparable time frames to their Western counterparts, but the order of the stages by which it develops follows a different path. The stages of ToM development are assessed using a scale developed by Wellman and Liu (2004). This scale measures mindreading ability through various tasks that reflect an increased understanding of others' minds. The scale consists of five stages. The first stage involves understanding the diversity of desires, where one recognizes that two people may have different desires regarding the same situation. For example, one child may want to play with a toy while another child may not. The second stage involves understanding the diversity of beliefs, where one understands that different people may have different beliefs about the same situation, without attributing any truth value to those beliefs. The third stage involves understanding access to knowledge. This stage involves recognizing that some individuals may possess information that others do not. For instance, if one child discovers the location of a hidden toy while another child remains unaware, the former will comprehend that only they know where the toy is. The fourth stage involves understanding false beliefs. At this stage, one can comprehend that a person may hold a false belief about a situation if they have not been exposed to the correct information. This is the case with the false belief test mentioned earlier. The fifth stage concerns understanding that people may hide or mask their real emotions. For example, a child might understand that someone might look happy externally even though they feel sad internally. Previous studies have shown that Iranian, Turkish, or Chinese children tend to develop an understanding of access to knowledge before an understanding of diversity of beliefs, while the opposite is true for Western children. This suggests that the timing and stages of development may vary depending on individual learning processes and experiences.

Although there are variations, it is well-documented that children in Western societies, who are the primary subjects of many studies, begin to recognize intentionality at the end of their first year of life. During this stage, children begin to perceive themselves and others as having a mind, which marks the end of the prelinguistic stage. An example of this cognitive development in infants is their ability to follow an adult's gaze. Infants follow an adult's gaze regardless of whether it is obstructed or not. Between 12 and 18 months of age, infants learn to distinguish between objects that obstruct or do not obstruct a subject's vision. For example, they learn to differentiate between an obscuring blindfold and a transparent one. As a result, they stop following an adult's gaze if it is obstructed (Meltzoff, 2008). Development continues to mature around 4 to 6 years of age, at which point children develop the concept of false belief (Johnson & Wellman, 1982; Richert & Harris, 2006). Cognitive abilities reach an almost complete developmental stage only around 10-12 years of age. This is accompanied by the understanding of concepts that do not apply to human and non-human animals or objective states of affairs, such as omniscience, the afterlife, and souls (Barrett et al., 2001).

As psychologist Henry Wellman (2017) points out, Current research on infants, although abundant, still has gaps. A common problem is that many studies highlight what infants can do, but do not delineate the limits of their abilities. For example, while they may demonstrate an apparent understanding of "false belief" in certain contexts, they may not demonstrate the same understanding in slightly modified situations. Studies such as those of Sodian and Thoermer highlight such inconsistencies: in some conditions, infants seem to understand the relationship between seeing and knowing, while in others they do not. A particular example is the "true belief after delay" condition, in which infants do not recognize an agent's "true belief" despite the agent having seen all relevant events. This suggests that infants' ability to interpret the beliefs of others is limited and influenced by context-specific variables. Current research focuses primarily on infants 15 to 18 months old, leaving a gap in our understanding of progressive development. In addition, recent research suggests that infants use statistical learning to understand social agents. This learning ability may explain part of their early development in theory of mind.

Psychologist Henry Wellman (2017) notes that although there is abundant research on infants, there are still gaps. Many studies focus on what infants can do, but fail to specify the limits of their abilities. For instance, while infants may appear to understand 'false belief' in certain contexts, they may not demonstrate the same understanding in slightly modified situations. Studies, such as those conducted by Sodian and Thoermer (2012), have highlighted inconsistencies in infants' understanding of the relationship between seeing and knowing. In some conditions, infants appear to understand this relationship, while in others, they do not. For instance, in the 'true belief after delay' condition, infants do not recognize an agent's 'true belief,' despite the agent having seen all relevant events. Research suggests that infants' ability to interpret the beliefs of others is limited and influenced by context-specific variables. The current focus of research is on infants aged 15 to 18 months, leaving a gap in our understanding of their progressive development. Furthermore, recent studies indicate that infants utilize statistical learning to comprehend social agents, which may contribute to their early development in theory of mind.

It is clear that this brief overview of the current debate leaves out many aspects. However, the information presented so far is sufficient to outline the main areas of discussion.

1.5 The Gap in the Theory of Mind Development Theories

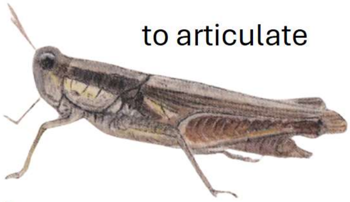
The preceding discussions have emphasized the significant link between language and ToM as highlighted by several philosophers and scientists. It is important to note that these thinkers have predominantly adopted a unimodal perspective on language when examining and supporting the connection between language and ToM. It is part of a longstanding debate in the history of philosophy and the study of animal behaviour to overlook the various modes of expression in humans and other animals. Aristotle, for example, argued that while animals can control the sound of words (ψοφος), they lack the ability to articulate speech (λογος) or engage in cognitive communication that expresses emotions such as pleasure or pain (φονη). Sextus Empiricus, who was both a theoretical source and polemical target for Descartes, argues that the difference between human beings and animals is not due to spoken speech, but rather to inner speech. Similarly, Gerauld de Cordemoy (1668) refers to the expressive manifestations of beasts as instinctual cries, a concept also explored by Lucretius and Montaigne. In the same period, Fabrici of Acquapendente wrote about identifying types of languages according to species in his work 'De brutorum loquela'. Regarding non-human animals, it is worth noting that even in humans, sign languages used by the deaf were not seriously considered until the 1960s. Looking back in time to Descartes, it was commonly believed that the deaf had lower cognitive abilities due to their lack of linguistic development. Language was not considered to have evolved from gestures and facial expressions until the pioneering ideas of Condillac and the work of Darwin (1882), followed by Romanes (1885), Mivart (1889), and Taylor, who took into account both sounds and gestures. This idea has been recently postulated by Corballis (2003). These innovative ideas remain unfortunately unheeded, possibly due to the strong impact of Noam Chomsky's work (1966) in identifying the mind-language set as specific to human beings. Chomsky's discontinuist approach made experimental research on the communicative abilities of animals scientifically uninteresting. In the last 20 years, a branch of studies related to a multimodal approach to communication has emerged. This approach enhances the entire expressive range of humans and other animals. Thanks to these studies, we have discovered new behavioural phenomena in nonhuman animals that were previously ignored. Due to the recentness of these studies or the inclination to maintain a discontinuous approach to language and communication research, these findings have not been taken into account by anyone, including the aforementioned philosophers. These findings have the potential to draw new conclusions about the functioning of the mind and to design new experimental protocols that could provide definitive evidence of the presence of ToM in non-linguistic animals. The following chapter will provide an in-depth analysis of these studies and their potential impact on the cognitive abilities of non-linguistic animals.

2 Animal language and communication studies: from unimodal approach to multimodality

The previous chapter briefly discussed the metaphysical aspects (which will be further discussed in Chapter 5) of ToM, in addition to its primarily epistemological perspective. This approach revealed a historically established dependence and co-evolution between language and ToM. The existence of natural languages in adult humans allows for accurate inferences of ToM from an epistemological point of view. In contrast, the absence of language in prelinguistic infants and nonhuman animals presents significant challenges to detecting ToM. Metaphysically, the interaction between language and ToM is examined from three distinct perspectives. Firstly, some theorists argue that language is a prerequisite for the development of ToM, suggesting that linguistic abilities enable the cognitive processes underlying ToM. Secondly, there is an opposing perspective that holds ToM as fundamental to language acquisition. This perspective argues that understanding the mental states of others is essential for language development. Finally, a third perspective challenges both dependency theories by proposing that ToM abilities can exist independently of language. This admits the possibility for nonhuman animals to engage in such cognitive activities despite not having language. ToM research often employs an unimodal approach to language and communication, which can introduce bias, especially in studies comparing cognitive differences between humans and other animals. This bias affects research in two main ways: it leads to the design of experiments that do not take into account the different perceptual and expressive abilities of animals, limiting our understanding of cognitive abilities in different species. An animal may not display a particular cognitive ability not because it lacks it, but because the experimental setting does not align with its natural communicative and perceptual patterns. Additionally, researchers tend to attribute distinctive features to human language from a metaphysical perspective, crediting these features with fostering unique cognitive abilities such as ToM. However, such abilities may exist in various forms and to varying degrees in other animals as well. This consideration highlights features like recursiveness and compositionality, which were once believed to be unique to human language but may not be solely responsible for the cognitive abilities we observe. These features, including recursiveness and compositionality, may not only be expressed through the vocalizations of non-linguistic animals but may also be part of a larger multimodal communication system that includes vocalizations, gestures, and facial expressions. Research suggests that communication in non-linguistic animals and humans is inherently multimodal and uses various signals through different sensory channels. This complexity of animal communication systems has been highlighted by several studies (Otovic & Partan, 2009; Partan, 2013; Taglialatela, Russell, Pope et al., 2015; Frölich & van Shaik, 2018). However, it has often been overlooked in the history of cognitive science. Recognizing the multimodal nature of communication systems is crucial for understanding animal and human linguistic abilities. The study of multimodality was first postulated by Darwin, and until the 19th century, certain notions could not be integrated into the study of animal behaviour or cognition. Only in the last 10 to 20 years has this research gained notoriety. The time frame for their results to find application in other fields of research, such as ToM, is too short. Considering the results of multimodality studies, what conclusions can we draw about the cognitive abilities of animals? These studies provide insight into the animal world. This chapter follows the structure of the previous one and offers a brief critical summary of the history of animal language and communication studies. It highlights how research on multimodality can aid in clarifying certain aspects of ToM. It points out how some characteristics of multimodality can be used to identify the presence of ToM and its level of development. Section 2.1 briefly discusses the historical unimodal approach that has accompanied this branch of study, from its origins to the earliest concepts of multimodality. Section 2.2 summarizes the history of how multimodality studies in humans and other animals originated and evolved. Section 2.3 analyzes the different methods of classifying multimodal signals and their critical issues. In Sections 2.4 and 2.5, I demonstrate the functional features of multimodal communication for investigating ToM in non-linguistic animals and for metaphysical postulations about the nature of the mind.

2.1 Brief history of the unimodal approach

The study of animal language and communication has its origins in the philosophical and cultural context of the fifth, fourth, and third centuries BCE. The Sophists were the first to identify lógos[20] as what sets humans apart from other animals. One of the earliest distinguishing features of human language from other animal communication systems is the ability to articulate sounds. Plato, Xenophon, and Hippocrates propose a distinction between two types of vocalizations: (i) simple emission of vocalizations (phoné), which is typical of animals or some humans, such as the deaf, and (ii) the ability to articulate them by means of the tongue (diálexis), which is exclusive to humans. Aristotle further develops this distinction by comparing the physiological and anatomical differences between humans and other animals. In his comparative analysis of various animal species, he proposes a further distinction by adding (i) the emission of vocalizations, (ii) their ability to articulate them, and (iii) the simple production of sounds (psóphos). For instance, many insect species, such as crickets and grasshoppers, produce sounds by rubbing their wings or legs against each other. Similarly, some birds, such as woodpeckers, produce sounds by beating their wings or drumming on wood with their beaks. The respiratory system facilitates the transition from simple sound production to actual vocalizations. Aristotle identifies this system as including organs such as the lungs and physiological structures such as the larynx, pharynx, and trachea, which are typical of mammals[21]. Additionally, the combination of the tongue, lips, and teeth enables the transition from basic sounds (phoné) to articulate speech (diálexis). Dogs provide an example of this limitation. Despite having a tongue and teeth, they cannot produce sounds with the same complexity as humans. The sounds of nonhuman animals are called agrámmatoi, meaning they do not have the minimal elements, such as phonemes, capable of compounding with each other to form complex sounds. Articulation and what will



---

[20] The Greek term 'logos' can be translated as 'word,' 'discourse,' or 'reason.' In ancient Greek philosophy, logos was considered the reason that determines the world and the law in which it is expressed. Logos was also associated with spoken language and rational thought. In this sense, logos can be seen as a synthesis of reason and language. Reason enables us to understand the world around us, while language enables us to communicate our ideas to others. Logos combines these two elements, enabling us to express our ideas clearly and rationally.

[21] Scientific knowledge in Aristotle's time was limited and differed from current knowledge, including that concerning animal respiration. Today, we know that there are many forms of alternative respiration, such as brachial respiration, which is typical of many aquatic animals like fish that use their gills to absorb oxygen from the water. Cutaneous respiration is practiced by some species of amphibians and insects that absorb oxygen through the skin. Insects respire through a system of tubes called tracheae, which is known as tracheal respiration. Other forms of respiratory apparatus have also evolved based on specific environmental requirements. Aristotle mistakenly attributed the ability to emit vocalizations to the respiratory apparatus of mammals alone due to a lack of knowledge.

later be referred to as the principle of compositionality are the characteristics that distinguish language from other forms of communication. These traits have been identified in recent times. Compositionality is often associated with the concept of recursiveness.

According to Sextus Empiricus[22], who is responsible for much of the evidence of the history of thought, the Stoics introduced a division of the concept of lógos. On one hand, there is lógos prophorikós, the proffered lógos, which is spoken language that manifests itself externally. (ii) On the other hand, we have the lógos endiáthetos, which is the cognitive counterpart of our external manifestations, or, to put it another way, the ability to think. According to the Stoics, what distinguishes humans from other animals is not speech (lógos prophorikós), but internal reasoning (lógos endiáthetos). This distinction was made in light of the imitative abilities of some birds, such as parrots, which can imitate human language[23]. Consequently, it was necessary to differentiate between similar articulatory and compositional abilities in nonhuman animals, without denying their existence. Therefore, if we interpret certain sounds in humans as manifestations of cognitive processes and inner reasoning, with the distinction of the two logos, the same sounds produced by nonhuman animals can be interpreted as instinctive or mechanical responses, not necessarily indicative of rational thought. Throughout history, many philosophers have defended the rationality of nonhuman animals, while others have denied it. However, this text will not focus on the issue of rationality, but rather on the historical approach to the study of language and communication. Before we move on to more recent times, let me quote another name from the distant past: Titus Lucretius Carus. In De rerum natura, Lucretius addresses the origin of language and argues that it has natural origins[24]. The

---

[22] Sextus was also the theoretical source and polemical target of Descartes (citation XXX).

[23] Aristotle discussed the imitative abilities of various birds, including their ability to imitate human speech and sounds. This observation is evident in his works on natural history, where he explores the behaviours and characteristics of animals, including their communicative abilities.

[24] In ancient Greece, two fundamental theories on the origin of language emerged. The first theory, known as Naturalism, posited that language had natural roots, reflecting the intrinsic essence of things. This theory was advocated by philosophers such as Heraclitus, Parmenides, and Democritus. In contrast, the second theory, called Conventionalism, argued that language was a human elaboration lacking an intrinsic connection to the objects or ideas represented. This theory was promoted by thinkers such as Protagoras and Gorgias. Plato, an ancient Greek philosopher, rejected both of these theories. He believed that language was a tool for achieving truth and knowledge, linking it to abstract and ideal ideas. Epicurus, the founder of Epicureanism, contributed to this discussion by proposing that language had natural origins but later developed into a conventional phase based on common linguistic rules. Lucretius was later influenced by Epicurus' writings.

need to produce different sounds arises from the fundamental desire to categorize what is around us. An example of this need, at least in human beings, is evident in infants during the prelinguistic age[25]. Infants use gestures to compensate for the later development of language. This need is not unique to humans, as other animals are also capable of emitting different vocalizations in response to different situations. For example, an animal in an aggressive state emits different noises than when it is in a quiet state, such as when caring for its young. Lucretius recognizes that animals express emotions, which is evidence of their fundamental need to categorize certain states of affairs. In addition to lógos prophorikós, Lucretius also sees this as an element of continuity between human beings and other animals.

Lucretius and other philosophers have referred to a prelinguistic phase where gestures played a role similar to that of language. Plato, in the Cratylus, reports that deaf people try to express their thoughts through gestures, movements of the head, and other parts of the body, suggesting the existence of a gestural communication system or an actual proto-sign language. Aristotle, on the contrary, believes that only individuals who speak a language can both express and possess thoughts. He asserts that the ability to hear language is the sole means of acquiring knowledge, rendering attempts to educate the deaf both impossible and futile. The matter at hand extends beyond the historical recognition of nonverbal communication, which has been minimal. It is not just about the extent to which scholars believe that these alternative methods of communication facilitate or limit the full expression and development of thought. The crucial issue is the depth and breadth of investigation that these forms of nonverbal communication have received. Gestures were initially recognized as a component of human communication primarily and, until recently, only in relation to rhetoric (Allan, 2013). Quintilian's Istitutio Oratoria, written in the 1st century CE, is a primary source on the topic of public speaking. The author provides detailed instructions on how to use one's voice, body posture, and gaze, as well as how to manage clothing. Of particular note are his instructions on hand gestures, which can draw attention to specific points in a speech, complete the steps of a syllogism, and indicate when facts are being stated or a conclusion is being reached. The situation will remain largely unchanged throughout the Middle Ages. During this period, body movements and gestures are primarily discussed from an ethical

---

[25] The idea of prelinguistic age is not found in Lucretius' works but is a modern definition.

perspective. The way in which a person moves their body and employs their hands during speech is considered an indicator of piety and goodness of heart, rather than cynicism or malice. Descriptions of how to use gestures during religious rites or prayers can be found. However, while there are books that explain how to use body language and gestures during a sacred speech or ritual, gestures are not typically a topic of academic inquiry. Additionally, the idea that other animals can also use gestures to communicate is not widely considered. As a result, comparative studies on the different communicative systems of humans and other animals typically only take vocalizations into consideration. Comparative studies often lead to the forced conclusion that humans are inevitably superior to all other animals. Human communication is often attributed with unique characteristics such as articulation, composition, and recursiveness, which may seem exclusive to humans if only vocalizations are considered. However, a multimodal approach that examines these features across various communication channels, rather than within a single sensory modality, may provide evidence of these features in different forms and developmental stages in other animal species. This perspective shifts from the traditional unimodal approach to a more inclusive understanding that recognizes the complexity of animal communication. The exclusive focus on vocalization as the sole significant means of communication has resulted in a biased perception of the role of language in cognitive development. This approach often overlooks other forms of communication, such as gestures and nonverbal signals, in the context of human and animal cognitive development. The limitation has resulted in a distorted view of the correlation between language development and cognitive development. This has reduced our understanding of the cognitive abilities of individuals with language disorders and nonhuman individuals. It has also contributed to an anthropocentric and abilist view of human superiority in ToM.

During the Middle Ages, emphasis was placed on distinguishing between spoken language (lógos prophorikós) and internal reasoning (lógos endiáthetos), as introduced by the Stoics. This distinction has become central to discussions and research on language and communication in both humans and other animals. An example of this can be found in Philo Alexandrinus' De animalibus, where he reports a dialogue between himself and his nephew Alexander. The former denies the presence of both types of lógos in nonhuman animals, while the latter claims it. According to Philo Alexandrinus, the difference between the voice of humans and other animals lies not only in articulation and compositionality, but also in the fact that it is an inseparable expression of thought. Therefore, it is linked to a cognitive element that is absent in the sounds produced by animals. Since animals lack thought (known as lógos endiáthetos), they cannot possess lógos prophorikós either. The sounds they produce are simply mechanical reactions based on instinct. Plutarch also draws attention to the cognitive aspect, arguing in Bruta animalia that the vocalizations of talking birds are not mere imitations of human language, but rather the result of a genuine process of teaching and learning. Sextus Empiricus argues that the voices of animals are a clear sign of the presence of a logos endiáthetos, using the division between the two lógoi as a conceptual basis. He compares these voices to those of barbarian peoples, which are unintelligible to those who are not part of them, but still endowed with full expressive-communicative functionality. Porphyry argues for a reunion between lógos prophorikós and lógos endiáthetos, stating that thought and speech are inseparable in humans and animals, forming a single lógos. This argument is similar to that of barbarian languages[26]. Despite the advent of Christianity, the situation remains largely unchanged. Throughout the medieval period, the arguments for and against the existence of language, or at least a communication system in nonhuman animals, remained consistent. Authors such as Arnobius, Lucius Caecilius Firmianus Lactantius, and Gregory of Nyssa followed this approach, leading up to two fundamental texts of the medieval period: Albertus Magnus' De animalibus and St. Thomas Aquinas' Summa theologica[27]. Here, we will find explanations of the differences between phoné, diálexis, and psóphos, as well as descriptions of the respiratory systems of humans and other animals, as in Aristotle. However, attention is given to the different types of voices found in the animal world, and it is highlighted how they work in an integrated way with sight and hearing for proper communication. This is one of the earliest attempts to study language and communication in a multimodal way, although it is not yet a comprehensive analysis. Other references to non-human animal language are found in Dante Alighieri's De vulgari eloquentia and Lorenzo Valla's works. Dante denies the possibility of real language to non-human animals, arguing

---

[26] The ancient Greeks had a perception of foreign languages, particularly those spoken by peoples outside Greek culture, which they referred to as 'barbarian languages'. These languages, often of Germanic origin, were perceived as incomprehensible and confusing. The term 'barbarians', used by the ancient Greeks to describe these peoples, was onomatopoeic and meant 'stutterers'. The Greeks struggled to interpret and understand these languages, as they sounded like a confused babble to them.

[27] The significance of these two texts lies in their widespread use, rather than in any new argument they bring to the debate on animal communication systems.

that they are guided only by instinct and do not require speech. He notes that talking birds are only able to imitate the sounds produced by humans, regardless of whether they are composed of complete words or not. They mimic human behaviour without comprehending the underlying concepts. In Dialecticae Disputationes, Lorenzo Valla challenges the notion that non-human animals lack reason and only possess instinct. Valla argues that those who make this distinction are playing a semantic game. Valla suggests that the concept of 'áloga' attributed to non-human animals may have originally been based on their inability to communicate through speech (oratio). Later, the term was extended to include their rationality as well. It is important to note that the term 'oratio' primarily refers to speeches or orations that are forms of verbal expression. Again, throughout the Middle Ages, the study of language continued to focus primarily on the vocal and auditory components, both in humans and in other animals.

In the modern age, there was increased intercultural contact between Europeans and people from non-European territories, particularly in the Americas (New World). This led to a growing sensitivity towards linguistic diversity, specifically in relation to human beings. As a result, a perspective emerged that gestural language could be a universal means of communication. This concept was inspired by the experiences of explorers and missionaries who reported success in communicating through gestures. Bonifaccio (1616), in his treatise 'L'arte de' cenni', aimed to promote the 'mute eloquence' of bodily gestures as a possible key to recovering the natural language of the human body, which is considered a divine gift accessible to all individuals[28]. The goal was to overcome the barriers imposed by the complexities of spoken languages. Condillac (1746) proposed a hypothesis that in a situation where a boy and girl, without language skills, were lost in a desert environment, they could develop a communicative system based on natural actions. This 'langage d'action' could have

---

[28] In the seventeenth century, philosophers noted the communicative significance of body posture, suggesting that it could reveal individual intentions or emotions. However, the term 'nonverbal communication' was formally coined only with Jurgen Ruesch, a psychiatrist, and Weldon Kees, an author. Prof. Albert Mehrabian made a significant contribution to the study of this phenomenon in the 1960s. Mehrabian developed the 7-38-55 communication model, which suggests that the meaning of a message is conveyed 7 percent through words, 38 percent through tone of voice, and 55 percent through body language, highlighting the predominant importance of nonverbal communication. It is important to note that this pattern applies more to emotional communications. Prior to the studies of Efron and a few others, research on nonverbal communication was limited. It did not become preeminent until the second half of the 20th century when eminent scientists began to explore the study of nonverbal language in the context of social psychology.

been the basis for early forms of communication and linguistic expression. During the 18th century, authors such as Diderot and Rousseau discussed the linguistic potential of gestures. Concurrently, a pedagogical movement emerged with a greater emphasis on using gestures to connect objects and expressions in a direct way, rather than focusing exclusively on word analysis. This movement aimed to reform education. During this period, notable figures emerged, such as Conrad Amman (1692), who developed techniques for rehabilitating the deaf and mute using his oralist method, and Charles-Michel de l'Epée (1771), who gained notoriety for his use of a system of 'methodical signs' in teaching deaf individuals, attracting the attention of the Parisian elite[29].

In modern times, research has expanded to include not only vocal expression in humans but also gestural communication. However, the discourse on nonhuman animals took a step backward with the introduction of René Descartes into the debate. Following the work of Lorenzo Valla, Descartes presented his influential 'Discourse on Method' in 1637, which significantly impacted the conversation on animal cognition and communication. Descartes postulates that the human being is a synthesis of two substances: res extensa, the physical and material component that is responsive to the laws of nature, and res cogitans, the immaterial and thinking component. Non-human animals, on the other hand, are excluded from participation in res cogitans and are consequently devoid of soul or consciousness, and thus subject only to the mechanical laws of nature. Language is considered the external counterpart of reason, a capacity that animals lack. According to the Cartesian thesis, creativity is a key characteristic of rationality, as opposed to the mechanical actions of non-human animals or automatons. This thesis was later developed by Noam Chomsky in the 1960s, who argued that the ability to use language creatively is one of the qualitative differences between humans and non-human animals. Descartes was the first to break away from the concepts of lógos prophorikós and lógos endiáthetos. He identified creativity as the distinguishing characteristic of language, rather than compositionality and articulatory ability.

---

[29] The oralist method in deaf education emphasizes the importance of spoken language while excluding the use of Sign Language. It is based on the ideology that the deaf person must speak. This approach involves the exclusive use of oral language and lip-reading, banning the use of gestures, which are considered counterproductive to the development of spoken language. In contrast, methodical signs were introduced by Charles-Michel de L'Épée in French sign language to translate grammatical elements of French, such as conjunctions and prepositions, and facilitate learning by the deaf. This innovation led to the creation of the first National Institute for Deaf Children in Paris, marking an important development in deaf education.

In a letter addressed to Henry More in 1649, Descartes confirms the idea of a clear distinction between human beings and non-human animals. He clarifies that language should be exclusively understood as the capacity to communicate the contents produced by the activity of this rational component. Additionally, he emphasizes that we should not confuse language with vocal expression alone. The philosopher defines speech as a certain sign of thought hidden in the body and notes that everyone, even those without the tongue or organ of voice, make use of it. It is important to note that the term 'word' is used synonymously with 'language' in a broader sense than is often used in cognitive science today. Language, in this context, refers to the ability to communicate thoughts through gestures, expressions, or other symbols, regardless of the ability to vocalize sounds. Non-human animals do not possess language, not because they cannot speak like humans, as talking birds can, but because they lack the cognitive ability to form and communicate thoughts. In his work Discours physique de la parole (1668), Gerauld de Cordemoy, a staunch supporter of Cartesian philosophy, argues that the vocal expressions of animals, including parrots, are mechanical and obligatory in nature. He compares the vocalizations of a parrot to the echo of a mountain, suggesting that just as it cannot be said that rocks speak, it cannot be said that parrots do.

The shift from language as vocalizations to language as the ability to communicate thought in any form led to arguments for a multimodal methodology. Until then, vocalizations were the primary form of expression studied. Let's review the steps that led to these early postulations. In his 1641 work Meditationes de prima philosophia, Pierre Gassendi engages in a debate with Descartes and echoes the argument previously made by other authors about the potential for non-human animal communication systems to be considered languages. Just as we recognize the languages of other human populations as languages, even though we do not understand them, we should do the same with other animals. According to Gassendi, the evaluation of the rational and linguistic abilities of nonhuman animals should not be based on human parameters. Instead, he advocates recognizing that each species has a unique mode of expression adapted to its specific ecological context and communicative needs. This perspective emphasizes the significance of comprehending animal behaviour in their natural environments and adaptation strategies. Gassendi advocates for this approach while also acknowledging the creative character of articulate speech and voice in non-human animals. However, other forms of expression are still ignored. The anonymous author of Theophrastus Redivivus (1659) discusses the Stoic differentiation between lógos prophorikós and lógos endiáthetos, as well as the Aristotelian tripartition of phoné, diálexis, and psóphos. The author argues that nonhuman animal sounds are full of meaning, even though they may be incomprehensible to humans. These sounds are a vehicle for expressing various states of mind. The author also suggests that animals without phoné, diálexis, or psóphos may have other ways of communicating their passions. It is reasonable to assume that these animals have distinct and effective means of communication. This idea, which proposes the existence of various communicative pathways in animals, will be revisited later by Condillac. He is cited above for his langage d'action (1746, 1755, 1775), which clearly supports two theses. (i) Language serves to express one's needs, so it can be as extensive as the needs of the species[30]. (ii) Not only sounds, but also different body movements are endowed with meaning and capable of expressing the different needs one has. This is the first time that the idea that other animals besides humans can also communicate through gestures is explicitly considered. It is important to note that while the inclusion of gestures is a significant advancement towards a more balanced study of modes of expression, the study of animal language and communication has long remained predominantly unimodal. Typically, vocal and gestural signals are analyzed separately, lacking an integrated approach.

During the seventeenth and eighteenth centuries, debates about the cognitive abilities of non-human animals were more prevalent than those about their communicative systems. This period saw the creation of the Encyclopédie (1751-72) by Denis Diderot and Jean Baptiste le Rond D'Alembert. The discussion of non-human animals and the works of the various authors who frequented the same milieu largely repeats what was said in previous debates, with a few minor additions. These additions do not contribute much from a conceptual or argumentative standpoint, but rather focus on the understanding of language. One notable exception is Guillaume-Hyacinthe Bougeant, who, in his work Amusement philosophique sur le langage des bêtes (1739), acknowledges the existence of different forms of language in various animals. The author suggests that movements and expressions of living things have meaning and communicative function. The voice is also considered a form of expression, but the author acknowledges the possibility of alternative communication systems. For instance, the

---

[30] Condillac's work, similar to Gassendi's, suggests an early ecological approach to the study of animal behaviour. It implies that communication systems are influenced by environmental and species-specific factors.

author argues that insects have a language suited to their needs and preservation, which we could perceive if we had the appropriate organs. According to Georges Leroy (1768), language is present even in non-human animals and serves both as a means of expressing emotions and social cognition. Language is not only an expressive function but also a means of categorizing a certain state of affairs and agreeing upon them through conventions.

This perspective, which considers animal communication in broader and more functional terms, was further developed by later naturalists and philosophers. Johann Gottfried von Herder's 'Abhandlung über den Ursprung der Sprache' (1772) explores the relationship between language, thought, and human nature, implicitly extending these reflections to the animal kingdom. In his 'Einleitung zum Kawi-Werk' (1836), Alexander von Humboldt delves into the relationship between language, thought, and the environment, laying the foundation for later investigations into animal communication. Within the field of biology, Jean-Baptiste Lamarck's 'Philosophie Zoologique' (1809) and Charles Lyell's 'Principles of Geology' (1830-1833) have influenced the understanding of language as an evolutionary phenomenon[31]. With the arrival of Charles Darwin and the publication of 'The Expression of the Emotions in Man and Animals' (1872), these ideas found fertile ground for a new understanding of language and communication. In this work, Darwin examines a wide range of communicative behaviours, both verbal and nonverbal, in different species. He highlights the use of a combination of vocal, gestural, and physical signals by animals to communicate. The author observes that animals' communicative modes serve not only to express emotional states or immediate needs but also as manifestations of complex evolution and adaptation to the environment. The author's theories suggest that communication, both verbal and nonverbal, evolves as a function of environmental needs and pressures, paving the way for a more integrated and dynamic approach in the study of interspecies communication. Darwin is credited with the original insight into the multimodal approach. This approach was developed and used in the

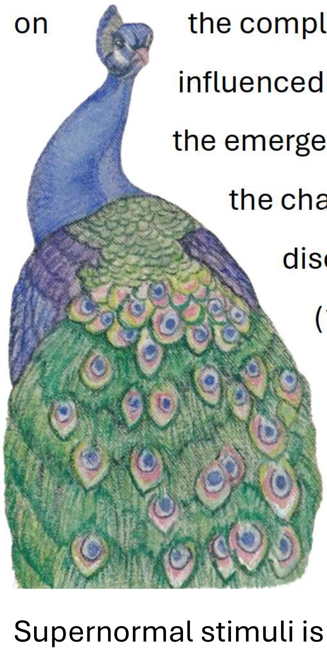
---

[31] Johann Gottfried von Herder, Alexander von Humboldt, Jean-Baptiste Lamarck, and Charles Lyell provided fundamental perspectives on the evolution of language in their research on communication and language. Herder's idea that animal sounds can be considered primitive forms of language contributed to a broader understanding of language's origin and evolution. Humboldt examined the influence of linguistic diversity on intellectual development, expanding the scope of communication studies. Lamarck and Lyell introduced evolutionary and geological concepts that can be applied to understand language as a phenomenon that changes and evolves over time. These contributions, although differing in their specific focus, have been crucial in shaping the current understanding of language as a dynamic and adaptive entity.

second half of the 20th century and the last years of the century. Many studies of animal communication have been conducted under the broader umbrella of the study of behaviour, starting with authors such as Leory and the rise of ethology. Although useful for a general understanding of animal behaviour, this categorization often fails to distinguish the unique and complex aspects of animal communication. Therefore, the emergence of multimodality studies represents a decisive breakthrough.

2.2 The emergence of multimodality

The multimodal approach to the study of language and communication has been established by two comprehensive reviews (Partan 2013; Higham & Hebets 2013) in a special issue of the journal Behavioural Ecology and Sociobiology. According to their respective authors, these reviews credit Darwin with the original insight of using a multimodal approach to study communicative systems and language, although he never explicitly used the term 'multimodal' in his writings. Instead, he concentrates on analyzing and describing signals sent through multiple sensory channels, particularly in relation to sexual selection. Darwin's focus on the complexity of communicative signals in natural selection significantly influenced subsequent research in biology and zoology. However, it was not until the emergence of ethology that studies of animal communication shed light on the characteristics of multimodality. Some of the founding fathers of the discipline, such as Tinbergen (1953, 1959), Marler (1965), and von Frish (1967), conducted these studies. Thus, animal behaviours such as supernormal stimuli and the waggle dance of the honey bee have been discovered. Additionally, studies have shown that vocalizations often accompany visual or gestural signals in different contexts, highlighting the multimodal nature of animal communication.



Supernormal stimuli is a phenomenon in which an animal's behavioural response to a signal is intensified by certain properties of the stimulus itself. For instance, if a bird is presented with a larger and more brightly colored egg, it may respond more strongly than to a normal-sized and colored egg. Male stickleback fish exhibit greater aggression towards other fish based on the vibrancy of the red color on their bellies. Similarly,

male peacocks use their colorful tail, specific calls, and pheromones to attract mates. A supernormal stimulus is an exaggerated version of a natural stimulus that elicits a stronger response from the recipient than the typical stimulus. The female peacock's mating response is intensified by the combined effect of these three different stimuli, making the male's display more effective than any single stimulus. This fragment exemplifies the principle of supernormal stimuli in animal behaviour, which can also be observed in human responses. It is important to note that this principle is not limited to animal behaviour. For instance, our responses to certain tastes and smells that signal essential nutrients, such as sweetness (indicating sugars) or umami (suggesting protein), can be amplified. By intensifying flavors and combining them with attractive visual and olfactory stimuli, junk foods effectively exploit biological responses, causing individuals to consume more of them than they might otherwise, despite awareness of the health implications.

The honeybee’s waggle dance is a prime example of multimodal communication. Bees combine visual signals, pheromones, and vibrations to convey reliable and accurate information about food location and quantity. The dance is a complex and highly structured form of communication that consists of two main types: the circular dance and the figure-eight dance. For instance, when the food source is within 50 meters of the hive, the bee performs a circular dance. However, if the distance is greater, the dance takes on the shape of an 'eight'. If the bee moves upward from the circle, it indicates that the new source is in the direction of the sun. Conversely, if it moves downward, the other bees interpret it as a signal to go in the opposite direction. If a bee cuts a circle at an angle, the other bees understand that they should fly either to the right or left of the sun, depending on the angle it makes relative to an imaginary vertical line. The vibrations produced during the dance are directly proportional to the amount of available food. Pheromones, on the other hand, can either facilitate or impair the learning and memorization of these cues (Baracchi et al., 2020).

Just like bees, the phenomenon of associating one signal with another is a common occurrence in various animal species and observed in different contexts. This is also true for mate attraction. For instance, many birds, including the peacock, use a courtship display that combines visual and vocal signals to express their desire to find a mate. The male peacock displays its colorful tail in a wheel, showcasing iridescent blue and green feathers while emitting vocal calls. These calls can range from high-frequency cries to low-frequency vibrations (Dakin et al. 2016). The purpose of this display is to attract a mate or to defend territory. In contrast, the domestic cat may puff up its fur, round its back, and show its claws to appear larger and more threatening while making a series of sounds such as growling or blowing. Finally, it marks its territory by releasing pheromones from specialized glands through rubbing its head against objects or urinating in specific areas (Draastad et al. 2022). Wild rabbits, for example, raise their tails and expose the white side as a visual alarm signal while beating their hind legs on the ground when they perceive a threat. All of these signals work together to communicate clear and consistent messages in a multimodal way (Monclùs et al., 2009).

Research on nonhuman animal multimodal communication significantly paused after the publication of 'Cartesian Linguistics' by Noam Chomsky in 1966, as noted in Fouts (1997). Although early investigations were conducted, further research was halted due to Chomsky's influential work. In his book, Chomsky proposed a discontinuist approach, arguing that human language did not evolve from primitive forms of animal communication[32]. The period of stalemate observed by Fouts, with some exceptions (McGurk & MacDonald 1976; Smith 1979; Stratton 1997; Uetz & Stratton 1983; Tomasello et al. 1985), resulted from the discontinuist perspective and the consequent reduction in research funding. Renewed interest in the topic of animal communication did not occur until the turn of the century. With the renewed interest in behavioural ecology, studies of nonhuman animals have focused on the evolution of multiple traits within a single sensory channel, particularly in the context of sexual selection. Studies have shown that these ornaments are often used by females to assess the quality of potential mates (Møller & Pomiankowski 1993; Schluter & Price 1993; Iwasa & Pomiankowski 1994;

---

[32] According to Noam Chomsky's discontinuist approach, human language is a unique and distinctive phenomenon that did not gradually evolve from more primitive forms of animal communication. This theory posits that human language has no common ancestor with animal communication systems, but rather represents a qualitative leap in cognitive capacity. Chomsky introduced the idea that humans possess an innate biological predisposition for language acquisition, known as the Language Acquisition Device (LAD). This device would allow humans to develop language ability based on a universal set of grammatical rules, known as 'universal grammar.'

Johnstone 1995, 1996; Rowe 1999). For instance, researchers attempt to understand why birds have different sexual ornaments[33]. Research involves not only the ability to produce and understand signals transmitted over different sensory channels for information exchange but also the effects that the transmission of multimodal signals has on the receiver. As we have seen, animals often produce and respond to signals composed of multiple components. This is an area of increasing investigation, as noted by Partan and Marler (1999) and Rowe and Guilford (1999). For example, the aforementioned birds frequently display sexual behaviours that combine both visual and acoustic elements in an elaborate manner. One possible explanation for the evolution of these complex signals is that they provide more reliable information to receivers than their unimodal counterparts by facilitating the entire communication process. It is worth noting that, in addition to these nonhuman animal developments, there is also a renewed interest in the study of human language. Currently, researchers are developing theories about the origin of language from gestural forms of

---

[33] Sexual ornaments are physical features that birds use to attract a mate. These features can include colorful feathers, long tails, and crests. Theories about sexual ornaments can be grouped into three categories. The first is the 'multiple message hypothesis,' which suggests that each ornamental feature conveys a specific message about the bird's qualities. For example, one type of ornament might indicate good health, while another might indicate physical strength. The second is the 'redundant signal hypothesis,' which proposes that each ornament provides part of the overall information about the bird's condition, with the set of ornaments providing a more complete picture of the bird's overall condition. Three theories have been proposed to explain the function of bird ornaments. The second is the 'redundant signal hypothesis,' which proposes that each ornament provides part of the overall information about the bird's condition, with the set of ornaments providing a more complete picture of the bird's overall condition. The first is the 'honest signal hypothesis,' which suggests that ornaments provide accurate information about the bird's overall condition. The third theory, the 'unreliable signal hypothesis,' postulates that some ornaments may not faithfully or accurately reflect the bird's overall condition. This hypothesis suggests that ornaments are retained within species not because they accurately convey information about health or strength, but because their production is not energetically costly. Therefore, these ornaments may persist throughout evolution even if they do not offer significant advantages in terms of sexual selection, simply because they do not require a high cost for the individuals who display them.

communication. Comparative research is also being conducted between humans and non-human primates, which contrasts with Chomsky's discontinuist approach.

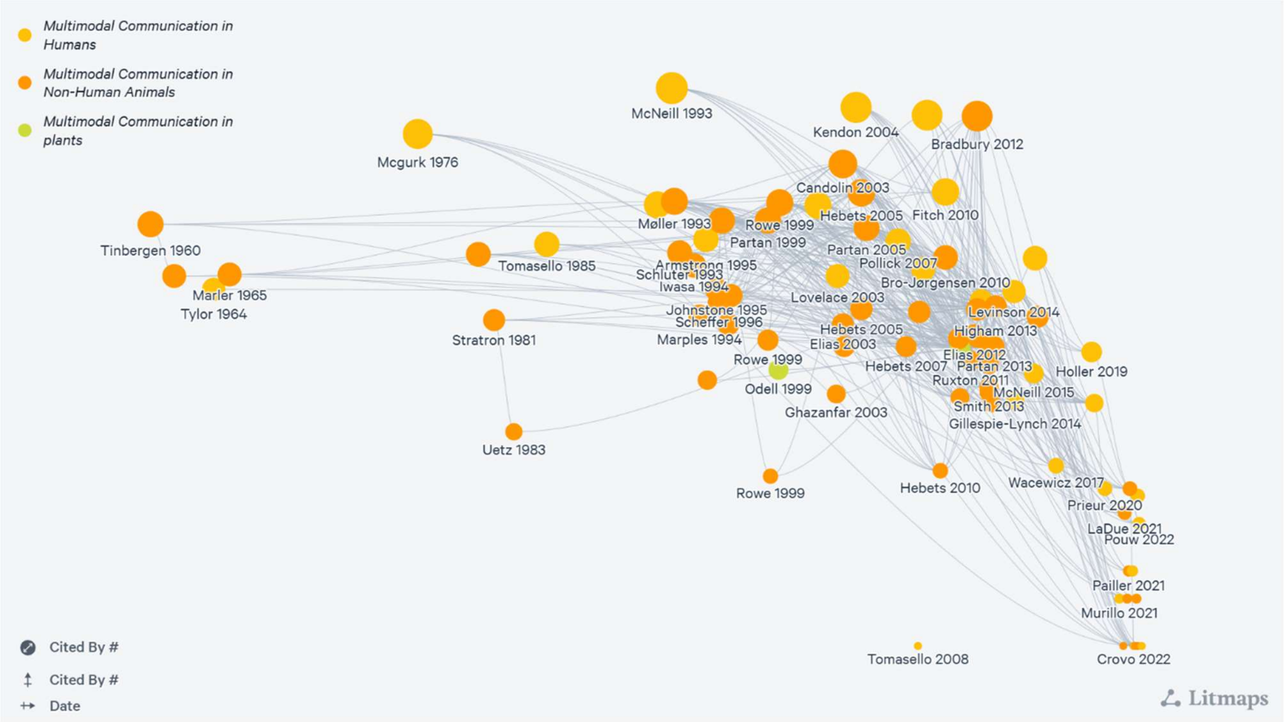
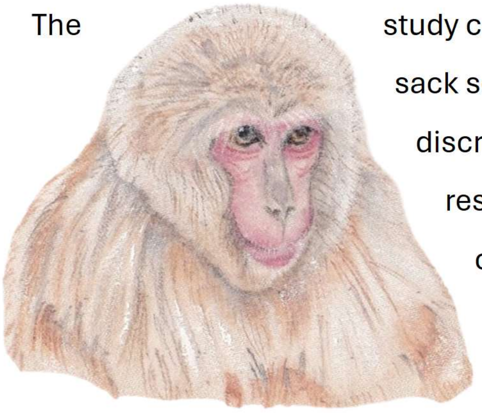


Figure 1 displays the frequency of publication of articles related to multimodal language and their impact in terms of citations. The diagram was produced using Litmaps software.

Rowe and Guilford (1996) in Nature and Partan and Marler (1999) in Science were major publications that opened up new research horizons. The study conducted by Rowe and Guilford investigates the nature and effectiveness of signals composed of multiple elements in animal species. The authors specifically examine how certain toxic insects use a combination of odors and bright coloration to warn potential predators of their danger. The experiment conducted by Rowe and Guilford focused on the use of pyrazine odor and flashy colorations, specifically red and yellow. They observed that the presence of pyrazine odor significantly increases aversion of predators, in this case, birds, to these colors. This effect was not observed when the odor was absent. Additionally, the researchers discovered that pyrazine alone does not elicit an aversive response and has no impact on the perception of non-hazardous colors, such as green. The main conclusion of their study is that the combination of odor and color, rather than each element individually, has a greater impact on predator psychology due to the interaction between the signal components. Partan and Marler's paper presents a novel framework for categorizing multimodal signals. The framework focuses on the responses of specific receivers, such as field ants, rhesus monkeys, and snapping shrimp, to isolated or combined signal components. The next section of the chapter (2.3) will provide a detailed exploration of this classification system.

Following pioneering studies by Rowe and Guilford (1996) and Partan and Marler (1999), researchers began conducting studies on multimodal communication in specific animals (Hebets 2005; Wilgers & Hebets 2011; Stasfstrom & Hebets 2013). For instance, Hebets examines the function of complex signals in wolf spider courtship. Using a novel experimental approach, Hebets tests the hypothesis that visual and vibratory signals in courtship function as mutual support under varying environmental conditions. Hebets' study found that the vibratory environment had a greater impact on mating frequency than the visual environment, which was unexpected. Additionally, he discovered that the presence of vibratory signals during courtship can affect females' receptivity to visual signals and even influence their visual attention.

Additional significant studies in the field of multimodal communication in nonhuman animals include those by Preininger et al. (2013) and Higham et al. (2013), following Hebets' contributions on wolf spiders. Preininger et al. (2013) researched ranids, commonly known as true frogs, to explore how environmental noise affects their multimodal courtship behaviours, which include both visual and acoustic signals. In their study of the small stream frog, the researchers found that while abiotic stream noise does not limit signal detection, the frogs must deal with interference and acoustic masking caused by the chorus of conspecifics. The researchers observed that multimodal signals, which combine acoustic and visual stimuli (such as an inflatable vocal sack), elicited a greater response in males than unimodal signals. The study conducted by Higham et al. (2013) suggests that the vocal sack serves as a visual signal that improves the detection and discrimination of acoustic signals in a noisy environment. The researchers investigated how different sensory modalities communicate various aspects of an individual's condition and how these signals are synthesized during mate selection in rhesus macaques. The study analyzed various factors such as sexual skin coloration, vocal cues, androgen levels, body



measurements, dominance hierarchy, and mating preferences of females. The results indicated that the most noticeable changes in sexual skin coloration involved brightness rather than hue, which did not consistently align with vocal signal characteristics. This discrepancy implies that sexual skin brightness and vocal signals might convey distinct types of information. This study is one of the first to incorporate multiple signal measurements in multiple modalities in nonhuman primates, which opens new perspectives in the analysis of animal communication.

Continuing the analysis of major research on multimodal communication in nonhuman animals, two important studies emerge: Munoz & Blumstein (2012) and Frohlich & van Schaik (2019). Munoz & Blumstein (2012) examine the importance of multisensory perception in the ecological context. They emphasize how animals use signals from different sensory channels, such as olfactory, sound, and visual, to reduce uncertainty about the environment. The study presents a theoretical framework based on psychophysics principles to comprehend the evolutionary importance of multimodal perception[34]. The authors emphasize that animals' perception of multisensory stimuli is influenced by ecological context and how anthropogenic activities can alter it, potentially leading to ecological consequences. Frohlich and van Schaik (2019) discuss multimodal communication in primate communication, emphasizing the importance of integrating the study of gestural and vocal communication. Their work reviews behavioural and neurobiological research, demonstrating that primate gestures and vocalizations share key language properties, including intentionality, reference, iconicity, and turn-taking. The study highlights the overlap in neurobiological mechanisms that produce gestures and vocalizations in primates, as well as their ontogenetic flexibility. These findings confirm that the origins of human language were multimodal. However, the authors note that in larger primates, gestures seem to have a more informative role in short-distance communication. In contrast, in human face-to-face interactions, the vocal channel predominates, suggesting an evolutionary shift from gestures to vocalizations as the primary mode of communication. These findings suggest new research directions for understanding

---

[34] Psychophysics is a branch of psychology that studies the relationship between measurable physical stimuli, such as light, touch, and sound, and the subjective response in terms of perceived intensity. It examines the relationship between psychological and physical or physiological phenomena. The term 'psychophysics' was coined by Gustav Theodor Fechner, who is recognized as the founder of this discipline, in 1860.

the origins of language and the link between complex communication, sociality, and the tendency toward cooperation.

It should be noted that this is not an exhaustive list of studies conducted since the turn of the century. Research on multimodality in nonhuman animals is vast and evolving, with many scholars contributing to a deeper understanding of how different animals use a range of sensory channels to communicate. Several studies have investigated the combined use of visual, acoustic, and olfactory signals and how these different forms of communication complement and influence animal behaviour and social interaction (Wong, 2014; Secondi et al., 2015). Previous studies have explored particular species, uncovering distinct patterns of multimodal communication. These range from the intricate courtship displays of certain bird species to the sophisticated social interactions of primates (Siniscalchi et al., 2018; Bensoussan et al., 2016; Taylor & Leung, 2020). Moreover, interdisciplinary research has been stimulated by the interest in multimodal communication, linking biology, psychology, neuroscience, and anthropology. This has offered new insights into the origins and evolution of complex communication. It has even been found that multimodal communication extends beyond the animal world, affecting plants as well (Schaefer & Ruxton, 2011; Opoku-Baah et al., 2021; Rosenthal & Ryan, 2000). Flowers are multisensory displays used by plants to influence pollinator behaviours. Leonard et al. (2011) provide a particularly fascinating example of this trend. Studies have shown that the complexity of floral signals, which combine visual, olfactory, and sometimes tactile elements, plays a crucial role in the reproductive success of plants. Plants seem to benefit from the integration of signals in multiple modalities, similar to animals.

In human studies, research on multimodality can be divided into two main categories: one focuses on multimodal communication in infants and neurotypical subjects, including those with speech disorders, while the other examines multimodal communication in adult humans. The former, which can be categorized into studies concerning non-linguistic animals, will be discussed in more detail in section 2.5. Indeed, the studies on infants and in neurotypical subjects deserve their own section because of the information they can provide us with about the phenomenon of multimodal communication. Of the studies concerning adult humans, however, I want to discuss them briefly. When trying to understand the role of multimodality in humans, studies fall into two categories. Those that investigate its nature at the neuronal level and those that investigate it from an evolutionary point of view. Let us focus on the latter for now. The former will be discussed later in Chapter 3. Three studies deserve special mention, particularly the first study by Hauser, Chomsky & Fitch (2002). This study marks the end of the discontinuist approach proposed by Chomsky in 1966 with Cartesian Linguistics. Although the rest of the scientific community still remains anchored in the philosopher's early formulations, this study is significant. Hauser, Chomsky, and Fitch propose a division of the language faculty into a Faculty of Language in the Broad sense (FLB) and a Faculty of Language in the Narrow sense (FLN) in their article. The authors challenge the notion that language is exclusive to humans. The FLB is responsible for perceiving and producing sounds, as well as forming and understanding concepts and intentions. The FLB and FLN are two distinct systems in the human brain. In contrast, FLN is the uniquely human ability to organize words and sentences in complex and recursive ways. In practice, FLN is similar to an advanced algorithm in the brain that enables humans to construct endless sentences by using grammatical rules to combine words in new and unique ways. Although FLN is a unique human capability, components of FLB can be observed, albeit in less developed forms, in other species as well. The article's three researchers explicitly state that if evidence of recursiveness is found in non-human animals, then the FLN ability would lose its human primacy.

The following studies are noteworthy: Tomasello's (2008) and Levinson and Holler's (2014). Tomasello presents a novel perspective on the evolution of human communication, emphasizing the role of social cooperation. This concept was first identified by Paul Grice. Tomasello posits that human communication is based on shared intentions and attentions, which evolved to facilitate cooperation and cultural life. Communication serves to aid, inform, and share ideas, thereby strengthening cultural ties. This need for cooperation influenced the development of grammatical structures in language. Tomasello suggests that human communication has its roots in natural gestures, such as pointing and pantomiming. Conventional communication evolved from these gestures, flanked by cultural learning that allowed the creation and transmission of common communicative conventions. According to Tomasello, the key difference between humans and other animals, particularly primates, is our advanced ability for 'mind-reading' (ToM) or intersubjectivity. This refers to our mutual understanding of intentions and mental states. In contrast to Tomasello, Levinson and Holler investigate the complex structure of human communication and point out how different communicative modalities (visual, auditory, tactile) each contribute uniquely to an integrated communicative system. They propose that the different layers of human communication, despite their different phylogenetic and evolutionary origins, function together harmoniously. The authors suggest that gestures and spoken language may have evolved together, indicating that they are part of a single communication system. This perspective allows for a more comprehensive study of language by placing it within a broader pragmatic, interactive, and multimodal context.

I have cited these three studies because they represent three important shifts in language and communication research. The first shift is from a discontinuist approach to an evolutionary perspective. The second shift is the idea that communication and language are based on a theory-of-mind foundation. The third concern is that communication and language are inherently multimodal and rely on an integrated communicative system that allows different sensory modalities to function harmoniously.

2.3 Classification methods and related issues

In the context of evolving perspectives on communication and language, discussions emerged regarding the definition of multimodal communication, signaling, and related categorization methods. The prevailing approach in the field often separates these two aspects: the definition of what constitutes multimodal communication and signaling, and the development of systems for categorizing them (Otovic & Partan, 2009; Ruxton & Schaefer, 2011; Higham & Hebets, 2013). This separation is likely due to practical considerations in research and analysis. The definition of multimodal signals is dependent on the conceptual frameworks used to classify them, making their description complicated. Currently, there is no unambiguous framework for their taxonomy. Additionally, due to the ambiguity surrounding technical terms in the field of animal communication, I aim to clarify my usage of these terms based on a specific tradition of studies. This will also define the meanings attributed to signals and multimodal communication. Similarly, this text aims to enumerate classification methods that, although incomplete and not integrated with each other, enable the identification of features that aid in comprehending the mechanism underlying certain cognitive phenomena.

Following Ruxton and Schaefer (2011) and Higham and Hebets (2013), a 'signal' is defined as any feature of an entity that can activate an observer's sensory system, potentially leading to a change in their behaviour or cognitive state. This influence could manifest as observable behavioural changes or as alterations within the brain that may not translate directly into visible actions. The definition of 'signal' is crucial in comprehending multimodal communication. By establishing the parameters of a signal, we can more effectively examine how various signals interact across sensory modalities to produce intricate communicative exchanges. This definition is based on the discussions of several authors, including Dawkins & Krebs (1978, 1984), Partan & Marler (2002), Shanker & King (2002), Rendall et al. (2009), and Owren et al. (2010). These authors argue that the traditional concept of communication as simply the transfer of information, as initially introduced by Shannon (1948) and further expanded by Weaver (1998), needs to be reconsidered. The authors suggest that communication involves more than just information transfer. Rather than solely transmitting information, communication should be viewed as a means of exerting influence. This perspective asserts that a signal's information content cannot be determined independently of the context and sensory abilities of the individual interpreting it. Therefore, greater emphasis is placed on the message's recipient rather than its author. In accordance with this approach, the traditional designations of 'sender' and 'receiver' have been redefined as 'informer' and 'perceiver,' highlighting the active role of the individual who interprets the signal. This change from information transfer to influence underscores the dynamic interplay between informer and perceiver, in which the significance and effect of a signal are jointly constructed based on the perceiver's context and sensory experiences.

Within this framework, a distinction is made between 'signals' and 'cues.' 'Signals' are defined as features that have evolved for their role in communication, according to authors such as Maynard Smith & Harper (2003), Schaefer & Ruxton (2011), and Higham & Hebets (2013). In contrast, 'cues' can still convey information and influence a potential 'perceiver,' but they lack the evolutionary adaptation for communication. "Cues" are environmental or incidental features that can provide information or influence the perceiver without being intentionally produced for communication[35]. To illustrate the distinction between signals and cues, Ruxton

---

[35] In this discussion, 'intentionality' refers to an individual's deliberate use of signals to communicate specific messages to others.

and Schaefer (2011) provide an example of a predator inadvertently stepping on a twig, producing a crackling sound that alerts the prey to potential danger. The prey, upon hearing this sound, will likely flee, interpreting it as an indication of the predator's presence. In this context, the crackling sound is a cue rather than a signal because it is an accidental byproduct of the predator's movement, not a behaviour intentionally produced to communicate. The sound results from the interaction between the predator and the environment, particularly its movement that causes the twig to break. This example highlights that environmental or accidental events, such as the sound of a twig breaking, can impact the behaviour of the recipient (in this case, the prey) without being specifically evolved for communicative purposes.

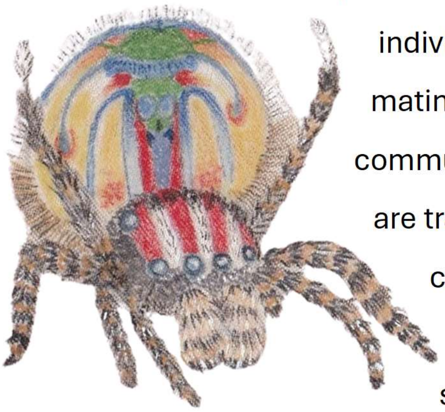
The distinction between signals and cues based on the concept of 'intentionality' is problematic due to the difficulty of attributing intentions to nonhuman animals. In this context, 'intentionality' refers to the purposeful or deliberate use of behaviour to communicate. The traditional approach in animal communication studies often links the evolutionary function of a behaviour to its intentional use as a communicative signal. However, this assumption can be misleading as it implies that behaviours with an evolutionary communicative function are always used intentionally. For instance, throwing a rock to catch attention exemplifies this issue. Although throwing has no evolutionary communicative roots and could be considered a cue, its deliberate use to attract attention turns it into a signal. This passage argues that the intention behind a behaviour can redefine it as a signal, separate from its evolutionary background. Therefore, it is not appropriate to automatically assume a link between the evolutionary function of a behaviour and its deliberate use in communication[36]. The argument challenges the definitions of 'signal' and 'cue' that rely on evolutionary function to infer intentionality, suggesting the need for a more nuanced approach in categorizing communicative behaviours. The wink gesture is an example

---

[36] In the study of animal communication, it is important to understand how signals develop from existing elements, known as 'protosignals.' Inferring intentionality from evolutionary function should be avoided. These may include body structures, organs, physiological processes, and ordinary behaviours. These protosignals originally lacked a signaling function. They undergo a process of ritualization, which simplifies and accentuates their components. This evolution suggests that while a signal may have a clear and specific role in animal communication, its origins and development may be less direct and more related to pre-existing functions and behaviours. Understanding this process is crucial for interpreting the intentionality and evolutionary function of signals in the animal kingdom today.

of how a behaviour, not originally evolved for communication, can become a meaningful communicative signal in human interactions. The wink can convey different meanings depending on the context, such as an invitation to flirt, a method of sharing a secret, or a signal of complicity in humor. This adaptability highlights that a behaviour or trait can gain communicative importance when used intentionally, regardless of its evolutionary origins. Consequently, the strict differentiation between signals and cues can be problematic, particularly when it is based on assumed intentionality. This differentiation may only be valid in hindsight, once intentionality can be inferred, if at all. Alternatively, the term 'signal' could be more accurately reserved for behaviours or phenomena that demonstrably influence the behaviour of others, without presupposing intentionality, thus eliminating the distinction between signals and cues. This approach recognizes the fluidity and complexity of communication, acknowledging that a behaviour's communicative value may emerge from its use and interpretation in specific contexts. The term 'signal' refers to any characteristic or behaviour of one entity that can influence the sensory system of another, potentially leading to a change in behaviour or cognitive state. This influence may result in observable changes or internal cognitive alterations that may not be directly visible. Signals may be intentional, where behaviour is used on purpose to convey information, or accidental, resulting from actions not specifically intended for communication but nevertheless capable of being interpreted by others.

Considering the discussion on communication flexibility and signal definition challenges, it is relevant to explore signal classification. Signals can be unique or composed of multiple distinct signals, known as multicomponent signals. Otovic and Partan (2009) introduced this concept. An example that illustrates this point is the human voice. It can convey complex information through variations in pitch, tone, and rhythm within the auditory channel. These components communicate not only the presence and identity of the individual but also crucial information regarding their health status, mating readiness, and even territorial demarcation. These types of communication are considered multimodal when their components are transmitted over different sensory channels. Multimodal communication is exemplified in the courtship behaviour of certain spider species. For instance, the male peacock spider performs a complex dance that includes intricate visual



movements, vibrations produced by tapping on the ground or leaves, and chemical signals to attract the female's attention. Higham and Hebets (2013) note that classifying signals can be complicated in some contexts because the definition based on physical properties excludes the sensory systems used by perceivers to detect the signal (Hebets 2011). Dambly-Chaudière et al. (2003) discuss the complexity of sound particle movements in marine environments and how they are perceived by different animals. While some species can detect these signals solely through hearing, others, such as certain aquatic animals equipped with neuromasts[37], can perceive them through other sensory channels. The question arises as to whether to classify this type of signaling as unimodal, involving only one sensory channel, or multimodal, involving multiple channels. Dambly-Chaudière and colleagues suggest that the latter classification, multimodal, is more appropriate, given the varied sensory reception in different species. I argue that the classification of a signal as multimodal or unimodal extends beyond its physical properties to include how it is interpreted within specific ecological contexts. The ecological context, which includes both the informer and receiver environments, plays a crucial role in determining the multimodality of a signal. Factors such as habitat acoustics, visual obstructions, and the sensory capabilities of the perceiver can influence signal transmission and perception. For instance, in a dense forest, visual signals may be less effective due to obstructions, emphasizing the role of acoustic signals. In contrast, in clear water, visual signals may be more important. Therefore, the ecological environment has a significant impact on the utility and interpretation of signals, which can alter their mode of perception.

If a signal is composite, its components can be considered fixed or free (Smith, 1977; Partan and Marler, 2005). A signal is considered fixed if it is produced as a direct consequence of the emission of another signal. For example, in human vocal production, lip movement must work in coordination with vocalizations (McGurk and MacDonald, 1976; Ghazanfar and Logothetis, 2003; Ghazanfar, 2013). In the case of frogs, the vocal sac swells when emitting a mating or advertisement call (Taylor et al. 2008; Grafe et al. 2012; Preininger et al. 2013). A signal is considered free if it is produced independently of another signal. Humans can produce free signals by gesticulating with their hands or winking during a speech, while pigeons can bow

---

[37] Neuromasts are specialized sensory cells capable of detecting mechanical stimuli, such as pressure or movement. They are found in aquatic vertebrates and make up the sensory system known as the lateral line.

and coo during courtship. However, this distinction is insufficient for signal analysis as it does not consider the possibility that a signal, even if considered free, may still be the result of an invariant response to a specific stimulus. I believe that two signals, which are not dependent on each other but are the product of an invariant response in certain contexts, should still be considered fixed and not free. Free signals, if they are the product of an invariant context, would lack the behavioural flexibility that characterizes them. This distinction is necessary to construct a conceptual framework for inferring cognitive abilities in non-linguistic animals, both human and nonhuman (Chapters 3 and 4)[38].

In summary, a signal occurs when an informant's aspect stimulates a perceiver's sensory system, exerting an influence. This influence can manifest in various forms, from observable behavioural changes, such as a response or action, to internal cognitive effects, such as the formation of a memory trace. Signals include anything capable of producing a change in the perceiver's state or behaviour, not just naturally selected behaviours. A signal can be singular or composite, meaning it can either stand alone or be a combination of multiple subsignals. When multiple signals are transmitted and detected through the same sensory channel, such as sound or sight, they are identified as multicomponent. In contrast, signals are considered multimodal when they are transmitted and received through different sensory channels, involving more than one type of sensory perception[39]. For instance, the perception of a signal on various sensory channels may involve both auditory and visual components. This can be exemplified by seeing a flashing light (visual channel) and hearing a siren (auditory channel). In animal communication, a bird may use both its colorful plumage (visual channel) and its unique song (auditory channel) during a courtship event. This approach enhances

---

[38] I do not assume the existence of these categories of signals in non-linguistic animals without empirical investigation, although it seems evident that they exist. Instead, I ask whether the existence of these two signal categories, within a well-structured theoretical framework, supports the attribution of higher cognitive abilities in non-linguistic animals. My answer is yes.

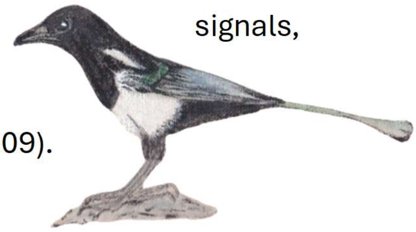
[39] In the context of the unified and composite signal mentioned in the text, it is important to distinguish between multimodal and bimodal communication. Bimodal communication involves the use of exactly two communicative modes. For example, a dog barking (acoustic communication) and wagging its tail (visual communication) is using bimodal communication, combining two types of signals to convey information. Multimodal communication refers to the use of more than two communicative modes. This approach is not limited to bimodal combinations, but can include a number of different modalities. For example, a bird might simultaneously employ singing (acoustic communication), feather movements (visual communication), and pheromone release (chemical communication). Thus, multimodal communication is a broad term that includes any communication that uses more than one modality. Bimodal communication is just one type of multimodal communication.

communication effectiveness by engaging multiple senses of the receiver, providing a more comprehensive set of information than a unimodal signal could convey. The study of multimodal signals has deepened the understanding of their use by animals in various contexts and their evolutionary development. To this end, several theoretical frameworks and classification methods have been developed. One approach emphasizes the fitness of the perceiver, exploring how multimodal signals influence and enhance the reproductive success or survival of nonhuman animals. In a survival context, a prey animal's ability to interpret and respond to multimodal alarm signals from conspecifics can significantly improve its chances of escaping predators. Selective pressures offer another perspective on multimodal signals, focusing on the environmental factors that determined their evolution. This approach examines how signals have adapted over time to enhance survival strategies, such as evading predators or communicating effectively with conspecifics. It is important to note that while previous discussions have highlighted the complexity of defining signals and cues, particularly with regard to intentionality, the recognition of evolutionary pressures does not contradict these points. In contrast, it enhances them by providing a wider context in which these signals function and develop, providing insights into their functional significance in the natural world. Additionally, there are criteria that examine the meaning of transmitted signals. In such cases, an effort is made to deduce the actual information conveyed through multimodal signals. Among the various methods of analysis and classification, the one that I consider the most problematic is the interpretation of interspecies communication due to inherent difficulties in translation[40]. Another method involves studying signal interaction in specific social or environmental contexts to understand how they affect signal transmission effectiveness (Hebets & Papaj, 2005). As previously mentioned, the older method of

---

[40] When faced with the challenge of interpreting the meaning of multimodal signals, the issue of interspecies translation and interpretation arises. In this context, Quine's principle of untranslatability, also known as the 'indeterminacy of translation' in the philosophy of language, is a relevant concept. Quine argued that there is no clear-cut translation between two different languages, as demonstrated in his thought experiment of the 'rabbit gavagai problem.' In this scenario, an ethnologist meets an indigenous person and attempts to translate the word 'Gavagai' spoken by the indigenous person while pointing to a running rabbit. The ethnologist hypothesizes that 'Gavagai' means 'Rabbit,' but the word could have different interpretations such as 'food,' 'white fur,' or 'fast movement.' This principle can be applied to the study of multimodal signals in animals, presenting a similar challenge. Translating or interpreting animal signals into human terms can be challenging or even impossible. The meaning of these signals can vary depending on the context, the animal's past experiences, or other factors not immediately apparent to human observers. The interpretation of multimodal signals is complex and nuanced due to their untranslatability and the variety of possible interpretations. These signals can have a wide range of potential meanings and functions in different behavioural and environmental contexts.

classification and analysis is based on analyzing the different behavioural responses of perceivers (Johnstone 1996, Partan & Marler 1999a, 2005). For the purpose of this research, I will explore this method in detail as it is considered the most useful. In fact, an approach based on the perceiver's response is the most functional for understanding the mechanisms underlying certain cognitive processes and identifying behaviours that suggest their presence. This approach is linked to some kind of interaction, collaborative or competitive, between two individuals, and is therefore more suited to shed light on cognitive abilities, such as ToM, that are closely related to the social behaviour of animals.



The original method classified multimodal signals as either redundant or nonredundant based on the perceiver's responses[41]. A signal was decomposed into its components and transmitted separately to the perceiver. The perceiver's response to each isolated component was recorded. If the behavioural response was the same for each component, the signal was considered redundant. If the behavioural response had differed, the multimodal signal would be defined as non-redundant. For instance, the magpie (Pica pica) exhibits an exceptional capacity to communicate with its parents using both visual and auditory signals, particularly when it needs food. In this specific case, we notice that both visual signals, such as the opening of the beak, and auditory signals, such as a particular call, convey the same request for food (Otovic & Partan, 2009). Regardless of the sensory channel used for perception, both visual and auditory signals provoke the same reaction in the parent, resulting in the act of feeding the young. This behaviour of the magpie represents a classic example of redundant signals. In this dynamic, the different signals do not offer additional or complex information, but rather reinforce the same message through different channels. The redundancy of signals ensures that the request for food is clearly communicated and received, regardless of environmental conditions or the attention status of the parents. For example, when a rhesus macaque perceives an aggressive signal that is exclusively visual in nature, its response tends to be one

---

[41] Therefore, not all signal interactions can be categorized as either redundant or non-redundant. Rowe (1999) criticizes Partan and Marler's approach, arguing that it overlooks the possibility that some components of a multimodal signal may not have an immediate, measurable meaning or behavioural effect when presented alone. The concept of 'enhancement' suggests that additional components may improve the learning of a message without providing extra information. These non-informative components could increase the detectability and discriminability of informational components for the receiver, without producing a measurable behavioural response on their own.

of submission. The situation becomes complicated when vocalizations are added to the visual signal, as the macaque is more likely to respond with aggression instead of submitting (Otovic & Partan, 2009). This behaviour is clear evidence that macaque behaviour is not solely driven by the isolated content of a signal but is significantly influenced by the combination of visual and auditory signals. This is typical and predictable behaviour under these circumstances, providing clear evidence that we are in the presence of a nonredundant signal.

The classification method based on perceiver responses was subsequently revised and expanded by Partan and Marler in 1999a and 2005. Subcategories were created for redundant and nonredundant signals. The former can be further divided into equivalent, enhancing, and antagonistic, as described by Munoz and Blumstein in 2012[42]. The latter can be classified as independent, dominant, modulating, or emergent. A signal is considered redundant if the perceiver's response to the isolated components of the multimodal signal or to the entire multimodal signal is qualitatively and quantitatively similar.

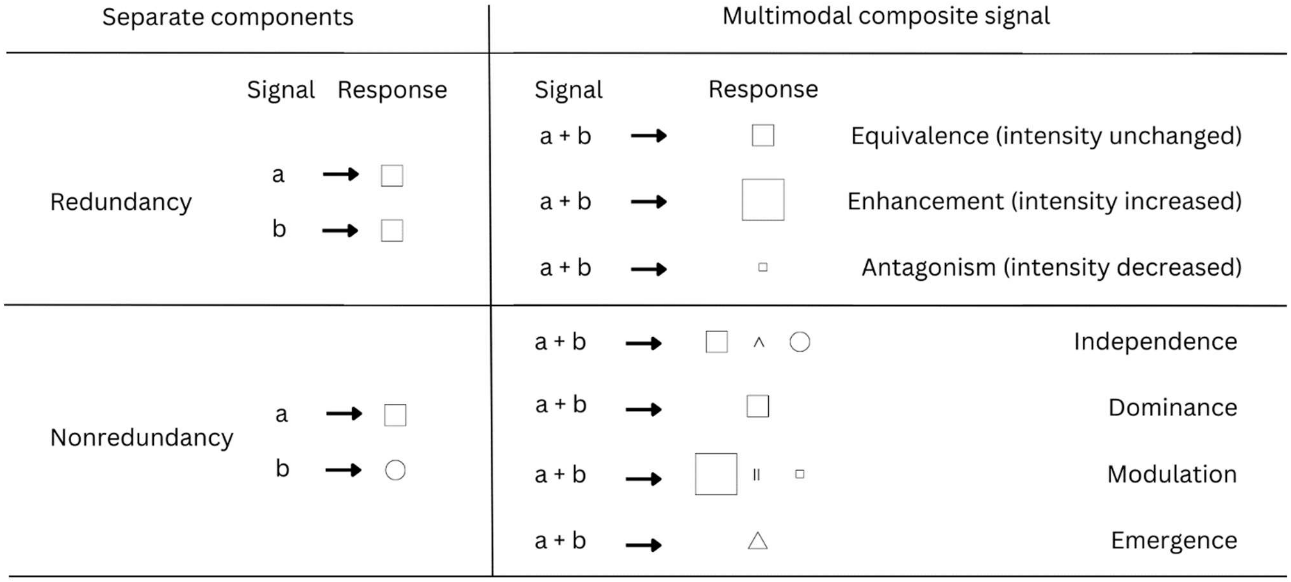


Figure 2 presents a modified version of the classification scheme of communicative signals based on perceptor response developed by Partan and Marler (1999), which includes the antagonists subcategory for redundant multimodal signals proposed by Munoz and Blumstein (2012).

---

[42] Munoz and Blumstein introduce a third category of redundant multimodal signals, which are predominantly found in anti-predator contexts. This category is supported by a narrow niche of studies, including Zuberbühler et al. (1999), Hazlett and McLay (2005), and Thompson et al. (2008).

Squirrels use both visual (tail movements) and acoustic (alarm calls) signals to indicate danger (Swaisgood et al. 1999, Rabin et al. 2006, Partan et al. 2010). When presented separately, both visual and acoustic signals induce a similar escape response in receivers, demonstrating their equivalence. In contrast, a signal is considered enhancing if the response obtained upon exposure to the complete multimodal signal is intensified compared to the responses observed with isolated exposure to the signal components. For instance, studies have shown that the alarm call of zebra finches, when accompanied by a visual signal such as wing flapping, causes a faster escape response, thus demonstrating the enhancing effect of combining signals. A signal is considered antagonistic if the response obtained from exposure to the entire multimodal signal is qualitatively identical but quantitatively inferior to the responses obtained from isolated exposure to the signal components. Prairie dogs use both visual (upright posture) and acoustic (alarm calls) signals to communicate danger. However, the combined use of these signals actually produces a less intense response in the receiver than the separate presentation of the signals. Nonredundant signals can be considered independent if the response to the exposure of the multimodal signal is only the sum of the responses obtained from the isolated exposure of its components. For instance, independent signals are observed in females of some spider species (Cross & Jackson, 2009; Hebets 2005), where pheromones released by females indicate their presence to males, signaling the possibility of mating. In parallel, the female produces vibration signals that guide males to her location. Each signal, although effective individually, maintains its functional independence when combined. 'Functional independence' in this case means that each signal, the pheromones and the web vibrations, retains its distinct purpose and effect. The pheromones communicate mating readiness, while the vibrations provide navigational assistance. When presented together, the component signals do not synergize to create a new or enhanced effect. Instead, each signal maintains its functional independence within the combined signal, eliciting the same response as when presented alone. A signal is considered dominant when one of its components dominates the response outcome over the others. The phenomenon of play behaviour in dogs is exemplified by the simultaneous bowing and growling exhibited during play (Bekoff, 1972). Despite growling being a typical threat signal, the visual signal of bowing takes precedence in this

context, transforming the interaction into an invitation to play. A signal is defined as modulating when the response to exposure of the multimodal signal is qualitatively similar but quantitatively different from the response to individual exposure of even one of the component signals. For instance, males of some shrimp species exhibit aggressive responses to visual signals alone, but modify this response in the presence of chemical signals from a female (Hughes, 1996). Finally, a signal is emergent when the observed response to exposure of the multimodal signal in its entirety is a novel response, qualitatively different from the responses obtained upon isolated exposure of the components of the multimodal signal. A classic example is the human perception of phonemes, in which the combination of misaligned auditory and visual stimuli produces the perception of a phoneme different from those presented individually.

Unfortunately, both the discussed classification method and other approaches often overlook the broader ecological context in which communicative exchanges occur. This neglect results in an incomplete understanding of the variability in the responses of different perceivers to the same signal. This variability can be attributed to differences in perceptual abilities (Ronald et al., 2012) or variations in individual experiences (Hebets & Vink, 2007). The ecological context is crucial as it determines the conditions under which communication takes place and affects how signals are interpreted by different receivers in the same environment. To comprehend these dynamics, it is necessary to consider the ecological factors that influence both signal transmission and the sensory abilities and experiences of receivers. Current approaches do not account for cases in which an individual may react differently to the same signal due to various boundary variables. These variables include an individual's current physiological state, previous experiences, or immediate environmental context. The term 'boundary variables' refers to situation - and individual - specific factors that can alter an individual's response to a signal. Ecological context, including factors such as habitat complexity, the presence of other species, and environmental conditions, can significantly impact how signals are perceived and interpreted. Also, it is worth noting that most research on multimodal communication has predominantly focused on a bimodal perspective, analyzing signals involving only two sensory modalities (Gleeson 1991; Bushmann 1999; Kamio et al. 2008). This approach may overlook the complexity and richness of multimodal signals that involve more than two sensory channels, which can limit our understanding of the full spectrum of animal communication.

Furthermore, since this field of study is relatively new, there are numerous unanswered questions that affect our comprehension of language and communication (Partan 2013). For instance, we have yet to understand the mechanisms of multisensory integration between different channels. The goal of this research is hindered by missing information, such as the identification of a common developmental trajectory of multimodal communication among taxa and whether it brings benefits in terms of individual survival and species reproduction. Additionally, it is unclear what cognitive capacities multimodal communication acts as an indicator of and whether ToM can be inferred as its prerequisite. Although there is a lack of information to facilitate this investigation, it is possible to explore the characteristics of multimodal communication based on current knowledge and draw conclusions from certain assumptions. Studies conducted on nonhuman animals have revealed peculiarities and phenomena related to communication (section 2.4) that suggest the presence of ToM in these animals. These findings demonstrate their ability to solve the epistemological problem (Chapters 3 and 4). Meanwhile, additional research conducted on humans, specifically on children with autism spectrum disorders and delayed language development (Meltzoff 1999, Murillo et al. 2021a) (section 2.5), indicates a potential directly proportional relationship between ToM and the capacity to produce and comprehend multimodal signals. The latter information is functional in solving the metaphysical problem of how ToM capacity works or develops (Chapters 3 and 5).

2.4 The multimodal shift phenomenon

Multimodal communication, both in humans and other animals, is characterized by a phenomenon known as the multimodal shift. This phenomenon, first anticipated by Marler (1967) and later canonized by Partan et al. (2010), involves a change in the sensory channel or configuration of channels used in a communicative context. This phenomenon occurs when a certain level of disturbance is surpassed, favoring less disturbed communication channels at that time (Brumm & Slabbekoorn 2005; Hebets & Papaj 2005; Partan & Marler 2005; van der Sluijs et al. 2011). A classic example for humans is attending a concert or being in a club. In these environments, background noise often exceeds the threshold for effective speech communication, leading people to rely on other sensory channels. In situations with high noise levels, individuals may resort to nonverbal communication such as gestures, facial expressions, or physical contact to convey their messages. For instance, a person may point to the bar to suggest going for a drink or use gestures to ask another person to follow them to a quieter place. This replaces the auditory (speech) channel with visual and tactile channels.

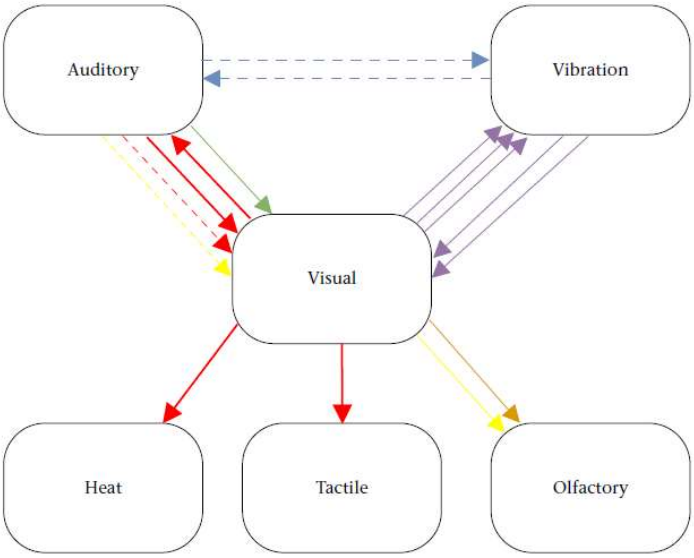


Figure 3 Reprinted from Partan 2017 show Sensory channels and direction of shift for the 16 multimodal shifts found in the literature. Each arrow represents one study, with most studies focusing on a single species, although some included multiple species. Each arrow represents one study, with solid arrows indicating immediate, context-dependent individual facultative shifts and dotted arrows indicating population-level shifts. The taxa are represented by different colors: mammals are red, birds are green, amphibians are yellow, fish are orange, insects are blue, and arachnids are purple.

Partan (2013) outlines the concept of multimodal shift, which is a type of phenotypic plasticity similar to sensory plasticity. In sensory plasticity, the enhancement of one communicative or sensory channel compensates for the deficit of another (Chapman et al., 2010; Berardi et al., 2000). This compensation often results from sensory deprivation during a critical developmental stage[43]. However, the temporal characteristics of phenotypic plasticity associated with multimodal shift differ from those of other neuroplasticity phenomena, which are generally permanent adaptations over an individual's lifetime. Multimodal shift represents a short-term behavioural adaptation. Therefore, after the disturbance that caused the sensory channel change ends, communication returns to the individual's preferred communication channel. For instance, in the case of a concert, after it has ended and the associated

---

[43] Chapman et al. (2010) discovered that fish raised in low-light conditions depend more on chemical cues than visual cues for foraging in later life.

disturbance has ceased, people resume communicating by talking to each other. They once again prefer the auditory channel over those used just moments before.

Multimodal shift, which is a transient phenomenon, can be influenced by the surrounding ecological environment[44], including both natural and artificial environments. This adaptability has been observed particularly in ex situ conservation settings[45]. Spiders were the first animals in which this phenomenon was observed. Researchers have found that several species of spiders can adapt their communicative behaviour according to the amount of light in their surroundings. For this investigation, researchers utilized a courtship arena specifically designed for observing pairs of spiders (male and female) during their courtship stages. The arena was equipped with high-speed cameras and accelerometers to capture their movements. The spiders were observed in both light and dark conditions. When there was little or no light, the spiders relied more on vibrational signals than visual ones, adapting dynamically to the environment (Scheffer et al. 1996; Taylor et al. 2005; Wilgers & Hebets 2011). Furthermore, chimpanzees have also been observed exhibiting the same behaviour. In an early observational study conducted by Tomasello et al. (1994), it was shown that chimpanzees use gestures to attract the attention of conspecifics who are looking at them, while seeking physical contact, such as touching or pulling objects, with those who have their backs turned or are looking away. Two more recent studies have also demonstrated that chimpanzees are capable of performing the multimodal shift even when communicating with humans rather than conspecifics (Leavens et al. 2010; Taglialatela et al. 2015). In the

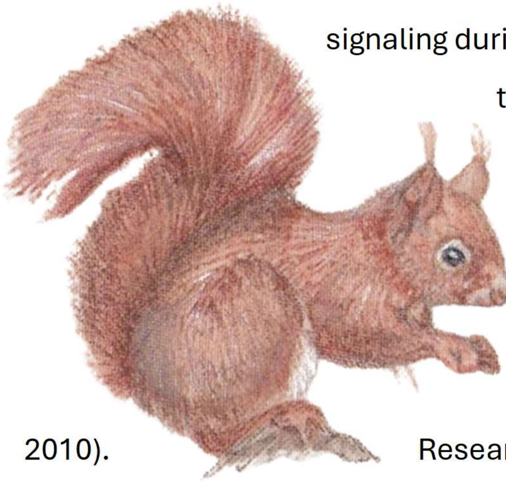
---

[44] It is important to note that an individual's ecological context includes not only their natural habitat but also man-made environments, such as urban areas or managed conservation spaces. These environments can significantly impact behaviour and communication strategies. The term 'ecological' refers to everything related to the interaction between living organisms and the environment in which they live. This fragment of text discusses the broad use of the term 'ecological' to refer to any context in which an individual interacts with their environment, including the natural, built, and social environment. It is important to note that this interpretation may not be universally agreed upon.

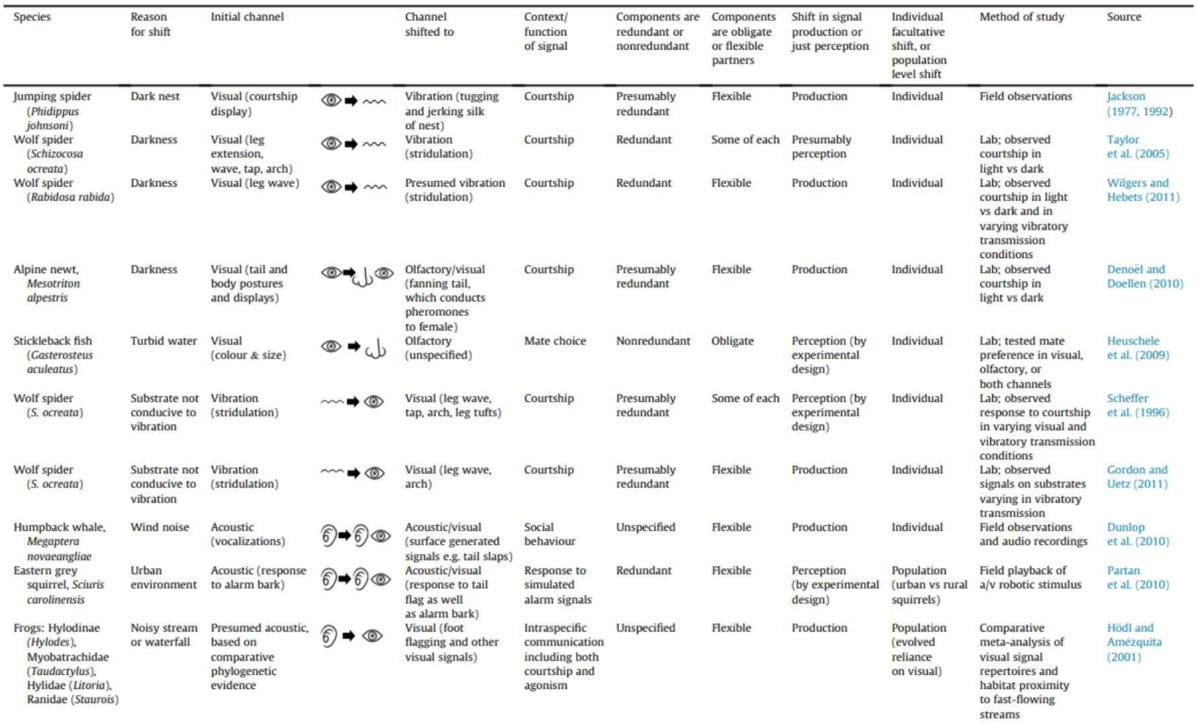
[45] Biodiversity conservation can be achieved through two main strategies: in situ and ex situ. In situ conservation aims to protect species in their natural environment, preserving the ecosystems and ecological processes that sustain their lives. This approach involves creating and managing protected areas, such as national parks and nature reserves, which are essential for maintaining ecological interactions and evolutionary processes. Ex situ conservation involves preserving species outside their natural habitat. This can be achieved through practices such as breeding animals in zoos or aquariums, growing plants in botanical gardens, and preserving genetic materials like seeds or embryos in germplasm banks. Ex situ conservation is especially valuable for highly threatened species or those that have already disappeared in the wild. It acts as a safety net and provides resources for possible reintroduction programs. These two approaches form an integrated system for conservation.

experiment conducted by Leavens et al., 110 captive chimpanzees were tested by being presented with food that they desired. During each trial, the experimenter faced one of the chimpanzees and then turned to face the other halfway through. The researchers evaluated how the chimpanzees adjusted their communicative behaviours based on the experimenter's orientation. They found that when the experimenter faced the chimpanzees, they predominantly used visual signals. In contrast, when the experimenter's back was turned, the chimpanzees relied more on auditory or tactile signals to capture the experimenter's attention.

Evidence of multimodal changes in animals within their natural habitats is relatively scarce. However, notable cases include studies of robins in urban environments (Fuller et al. 2007). The impact of the urban environment on their communication strategies has been documented. Robins are highly territorial birds dependent on vocal communication. However, noise pollution in cities interferes with the propagation of acoustic signals in the environment. This prompts individuals to alter their communication behaviour by visually signaling during the day and using vocalizations at night, in contrast to those living in non-urban areas. Similarly, gray squirrels living in urban environments have been observed relying more on visual tail wagging as an alarm signal than on auditory signals, compared to those living in rural areas (Partan et al., 2010). The most recent observed case in the wild pertains to humpback whales (Dunlop et al. 2010). Researchers discovered that the whales gradually transition from using primarily vocal communication to primarily visual communication, utilizing signals such as breaching and pectoral slapping. This multimodal shift occurs in response to noisy environmental conditions that disrupt vocal signals, specifically when wind speed and sea background noise levels increase. The researchers used a set of five hydrophones with a wide base to locate the sources of social sounds emitted by the whales. With this setup, the authors recorded 998 sounds from 49 groups of migrating humpback whales. This approach enabled them to precisely determine the origin of the sounds and analyze variations in the whales' communicative behaviour in response to their acoustic surroundings.

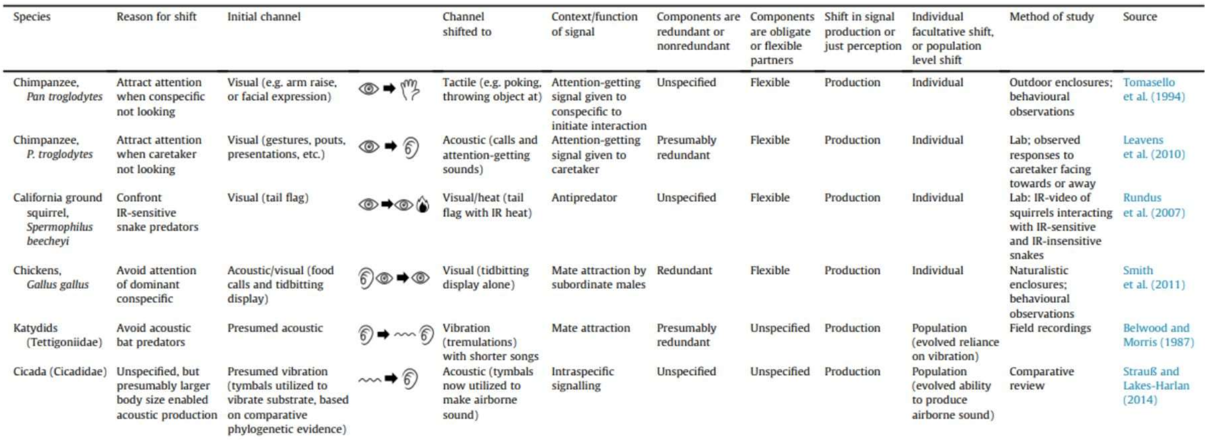


According to Partan's (2017) review, there are fourteen species involved in the multimodal shift phenomenon. The limited number of species is understandable, considering that this behaviour was only identified a little over a decade ago (Partan, 2013). It is noteworthy that this group includes taxa from mammals, birds, amphibians, fish, insects, and arachnids. The data indicate a potential spread of multimodal shift throughout the entire animal kingdom. This phenomenon appears to be associated with both abiotic environmental factors (63 percent) and biotic factors (36 percent), according to current knowledge. The observed phenomenon appears to be caused by either abiotic factors such as climate, soil, water, air, light, temperature, and pollution, or social or anti-predatory reasons towards conspecifics or other species. In some species, the multimodal shift is performed in a widespread and homogeneous way at the population level, as in the case of squirrels or robins mentioned earlier. All individuals of this species exhibit the multimodal shift under the same conditions and targets, such as visually warning a conspecific of danger in an urban and therefore audibly disturbed context. However, in most cases, the multimodal shift is an individual behaviour (75%). This means that not all individuals of a species necessarily display the same type or degree of multimodal shift in their communicative behaviour. Therefore, changes are specific and unique to individuals within a species. Differences in temperament, past experience, or physical condition play a major role in this regard. This variability makes understanding and observing multimodal shift even more difficult because it requires a detailed examination of the behaviour of individuals rather than species-level generalizations.



Species	Reason for shift	Initial channel		Channel shifted to	Context/ function of signal	Components are redundant or nonredundant	Components are obligate or flexible partners	Shift in signal production or just perception	Individual facultative shift, or population level shift	Method of study	Source
Jumping spider (*Phidippus johnsoni*)	Dark nest	Visual (courtship display)	👁 ➡ 〰	Vibration (tugging and jerking silk of nest)	Courtship	Presumably redundant	Flexible	Production	Individual	Field observations	Jackson (1977, 1992)
Wolf spider (*Schizocosa ocreata*)	Darkness	Visual (leg extension, wave, tap, arch)	👁 ➡ 〰	Vibration (stridulation)	Courtship	Redundant	Some of each	Presumably perception	Individual	Lab; observed courtship in light vs dark	Taylor et al. (2005)
Wolf spider (*Rabidosa rabida*)	Darkness	Visual (leg wave)	👁 ➡ 〰	Presumed vibration (stridulation)	Courtship	Redundant	Flexible	Production	Individual	Lab; observed courtship in light vs dark and in varying vibratory transmission conditions	Wilgers and Hebets (2011)
Alpine newt, *Mesotriton alpestris*	Darkness	Visual (tail and body postures and displays)	👁 ➡ 👃 👁	Olfactory/visual (fanning tail, which conducts pheromones to female)	Courtship	Presumably redundant	Flexible	Production	Individual	Lab; observed courtship in light vs dark	Denoël and Doellen (2010)
Stickleback fish (*Gasterosteus aculeatus*)	Turbid water	Visual (colour & size)	👁 ➡ 👃	Olfactory (unspecified)	Mate choice	Nonredundant	Obligate	Perception (by experimental design)	Individual	Lab; tested mate preference in visual, olfactory, or both channels	Heuschele et al. (2009)
Wolf spider (*S. ocreata*)	Substrate not conducive to vibration	Vibration (stridulation)	〰 ➡ 👁	Visual (leg wave, tap, arch, leg tufts)	Courtship	Presumably redundant	Some of each	Perception (by experimental design)	Individual	Lab; observed response to courtship in varying visual and vibratory transmission conditions	Scheffer et al. (1996)
Wolf spider (*S. ocreata*)	Substrate not conducive to vibration	Vibration (stridulation)	〰 ➡ 👁	Visual (leg wave, arch)	Courtship	Presumably redundant	Flexible	Production	Individual	Lab; observed signals on substrates varying in vibratory transmission	Gordon and Uetz (2011)
Humpback whale, *Megaptera novaeangliae*	Wind noise	Acoustic (vocalizations)	👂 ➡ 👂 👁	Acoustic/visual (surface generated signals e.g. tail slaps)	Social behaviour	Unspecified	Flexible	Production	Individual	Field observations and audio recordings	Dunlop et al. (2010)
Eastern grey squirrel, *Sciuris carolinensis*	Urban environment	Acoustic (response to alarm bark)	👂 ➡ 👂 👁	Acoustic/visual (response to tail flag as well as alarm bark)	Response to simulated alarm signals	Redundant	Flexible	Perception (by experimental design)	Population (urban vs rural squirrels)	Field playback of a/v robotic stimulus	Partan et al. (2010)
Frogs: Hylodinae (*Hylodes*), Myobatrachidae (*Taudactylus*), Hylidae (*Litoria*), Ranidae (*Staurois*)	Noisy stream or waterfall	Presumed acoustic, based on comparative phylogenetic evidence	👂 ➡ 👁	Visual (foot flagging and other visual signals)	Intraspecific communication including both courtship and agonism	Unspecified	Flexible	Production	Population (evolved reliance on visual)	Comparative meta-analysis of visual signal repertoires and habitat proximity to fast-flowing streams	Hödl and Amézquita (2001)

Figure 4.1 displays the species in which a multimodal shift was identified due to abiotic factors. The figure is adapted from Partan (2017).



Species	Reason for shift	Initial channel		Channel shifted to	Context/function of signal	Components are redundant or nonredundant	Components are obligate or flexible partners	Shift in signal production or just perception	Individual facultative shift, or population level shift	Method of study	Source
Chimpanzee, *Pan troglodytes*	Attract attention when conspecific not looking	Visual (e.g. arm raise, or facial expression)	👁 ➡ ✋	Tactile (e.g. poking, throwing object at)	Attention-getting signal given to conspecific to initiate interaction	Unspecified	Flexible	Production	Individual	Outdoor enclosures; behavioural observations	Tomasello et al. (1994)
Chimpanzee, *P. troglodytes*	Attract attention when caretaker not looking	Visual (gestures, pouts, presentations, etc.)	👁 ➡ 👂	Acoustic (calls and attention-getting sounds)	Attention-getting signal given to caretaker	Presumably redundant	Flexible	Production	Individual	Lab; observed responses to caretaker facing towards or away	Leavens et al. (2010)
California ground squirrel, *Spermophilus beecheyi*	Confront IR-sensitive snake predators	Visual (tail flag)	👁 ➡ 👁 🔥	Visual/heat (tail flag with IR heat)	Antipredator	Unspecified	Flexible	Production	Individual	Lab: IR-video of squirrels interacting with IR-sensitive and IR-insensitive snakes	Rundus et al. (2007)
Chickens, *Gallus gallus*	Avoid attention of dominant conspecific	Acoustic/visual (food calls and tidbitting display)	👂 👁 ➡ 👁	Visual (tidbitting display alone)	Mate attraction by subordinate males	Redundant	Flexible	Production	Individual	Naturalistic enclosures; behavioural observations	Smith et al. (2011)
Katydids (Tettigoniidae)	Avoid acoustic bat predators	Presumed acoustic	👂 ➡ 〰 👂	Vibration (tremulations) with shorter songs	Mate attraction	Presumably redundant	Unspecified	Production	Population (evolved reliance on vibration)	Field recordings	Belwood and Morris (1987)
Cicada (Cicadidae)	Unspecified, but presumably larger body size enabled acoustic production	Presumed vibration (tymbals utilized to vibrate substrate, based on comparative phylogenetic evidence)	〰 ➡ 👂	Acoustic (tymbals now utilized to make airborne sound)	Intraspecific signalling	Unspecified	Unspecified	Production	Population (evolved ability to produce airborne sound)	Comparative review	Strauß and Lakes-Harlan (2014)

Figure 4.2 shows the species in which a multimodal shift was identified due to biotic factors, such as social or antipredator behaviour. This figure is a reprint from Partan (2017).

The phenomenon of multimodal shift can be classified according to several characteristics (Partan 2017). The function of the signal is the first characteristic, which can be related to various behavioural aspects such as courtship, partner choice, social communication, alarm, attention, intraspecific communication, or for antipredatory reasons. The question at hand is whether the multimodal shift occurs in both signal production and perception (88%) or solely from a perceptual standpoint (12%). It is important to distinguish whether both the informant and the perceiver make the shift during communication or if the shift is made exclusively by the perceiver. For instance, in the case of squirrels, alarm signals that consist of both visual and audible components are not shifted by the informer. The perceiver is the only one who, when unable to listen undisturbed to the sound component, makes a perceptual shift by focusing mainly on the visual component. It is also important to consider whether the multimodal shift is based on redundant components (63%) or not (37%), and whether the relationship between components is free (69%) or fixed (31%). These classifications are based on the same ones used for multimodal signals explained earlier (2.3).

The detailed process of classifying multimodal shift is explained to consider a specific class of this phenomenon as a potential indicator of the presence and degree of development of ToM ability in non-linguistic animals. This consideration will be used to build the theoretical framework (Chapter 3) to address both the epistemological and metaphysical problem. Furthermore, through the use of the multimodal shift mechanism, I plan to create experimental protocols that can empirically support the theoretical arguments presented in Chapter 4. Before delving into the core of this research, it is essential to consider the findings of studies on multimodal communication in children with autism spectrum disorder and delayed language development. A theoretical framework that explains and identifies the mechanisms and degrees of ToM development in non-linguistic animals cannot solely rely on considerations from the world of nonhuman animals. It must also incorporate data gathered from research on humans, and harmonize them with the former.

2.5 Multimodality in pre-linguistic infant and children with ASD and LD

When examining research on human multimodal communication, two distinct types of studies are encountered, as outlined in section 2.2. One category deals specifically with infants, including those with speech disorders, and neurotypical subjects, and the other examines multimodal communication in adult humans. In section 2.2, studies related to multimodal communication in adult humans were briefly discussed. Now, let's focus on the first category of studies, which can be further divided into two groups: The first group of analyses, conducted since the 1980s, indirectly addresses the issue of multimodality in humans. Although the term 'multimodal' is not used, these analyses are interested in the interaction of tactile, auditory, and visual channels from both a productive and perceptual standpoint. Examples of such analyses include Meltzoff and Borton (1979), MacKain et al. (1983), Meltzoff and Gopnik (1993), Walton and Bower (1993), and Meltzoff (1999). The second group of studies comprises more recent research (Murillo et al. 2021a, 2021b; Valle 2022). These studies establish a connection between prelingual infants and children with autism spectrum disorder (ASD) and delayed language development (LD) and multimodal communication. We will further explore both groups of studies.

Studies in the first category have contributed to four key findings that have influenced the understanding of cognition in infants and young children. These findings include: the ability of infants to connect different sensory inputs, a phenomenon known as intermodal matching or cross-modal binding; the ability to mimic the lip and tongue movements of adults; the ability to store information long-term for later use in contexts other than learning; and, additionally, ToM development, which reaches a mature stage between the ages of 4 and 6 years, has been observed to occur as early as 18 months. The most significant study regarding multimodality is the one concerning cross-modal binding (Meltzoff & Borton, 1979). In the familiarization phase, researchers provided 29-day-old infants with the chance to explore an object at the tactile level. The experiment involved showing children two images, one of which corresponded to the object they had touched. The time spent observing the images of the two objects was then measured. Out of the 32 subjects tested, 24 fixated more on the image that corresponded to the object they had previously experienced. The experiment was repeated with another 32 subjects, and 22 of them stared longer at the image related to the object from the familiarization phase[46]. The experiment highlights how human beings, even in early stages

---

[46] Intermodal coordination or cross-modal binding differs from Jean Piaget's theory of sensory development in infants. Piaget believed that infants have separate and unconnected senses that gradually integrate over time. In contrast, cross-modal binding proposes that infants possess an innate ability to integrate information from different senses, such as sight and hearing, demonstrating a more advanced multisensory understanding than Piaget assumed. This concept challenges Piaget's idea that sensory and cognitive development in early life occur sequentially.

of development, can integrate information from different sensory inputs coherently. This is confirmed by later studies (Murillo 2021b).

This conclusion is consistent with recent research on children with ASD and LD. Specifically, Murillo et al.'s (2021a) study, although conducted on a small sample of subjects, indicates that children with these disorders have a lower ability to combine different communicative elements compared to children with typical development (TD). The study involves 22 participants. Half of the participants were children with autism spectrum disorder (ASD) aged 28 to 79 months, and the other half were typically developing (TD) children aged 12 to 30 months. The communicative productions were observed during semi-structured play sessions and recorded for later analysis. The children were paired based on the extent of their vocabulary[47]. The researchers used the Mann-Whitney U test[48] to analyze the differences in gesture production and vocal behaviours between the two groups. The study results indicate that during the initial stages of language development, there are no significant differences in the frequency of communicative behaviours between the ASD and TD groups. However, as language skills progress, TD children exhibit a higher frequency of communicative behaviours than ASD children[49]. This trend is not observed in children with ASD, suggesting differences in communicative development and potential challenges in cognitive flexibility[50]. Differences are

---

[47] Productive vocabulary extension refers to the number of words a child can actively use in communication. This is a measure of the child's ability to express themselves verbally.

[48] The Mann-Whitney U test, alternatively known as the Mann-Whitney-Wilcoxon test or the Wilcoxon rank-sum test, serves as a nonparametric statistical method. It is employed to assess the null hypothesis that two independent samples, X and Y, from different populations have equal distributions. Specifically, it tests whether the probability of a randomly selected value from the first sample (X) being greater than a randomly selected value from the second sample (Y) is the same as the probability of Y being greater than X. The Mann-Whitney U test is particularly useful in situations where the assumptions of normal distribution and equal variances (homoscedasticity) required by parametric tests, such as the two-sample independent t-test, are not met. It is applicable to both small and large sample sizes, making it a versatile tool in nonparametric statistical analysis, especially when dealing with skewed distributions or ordinal data.

[49] Gestures were classified according to criteria from previous studies. The categories included pointing, reaching, showing, giving, symbolic gestures, conventional gestures, signs of an augmentative communication system, instrumental use of an adult's body part, and other gestures that did not fit into the previous categories. Vocalizations were coded only when there was a second of silence or a turn of conversation between them, excluding shouting, laughter, and vegetative sounds. The study classified communicative behaviour into babbles, words, word combinations, echoes, and others. Additionally, making eye contact or looking at the adult's face was coded as part of communicative behaviour. The study found that a behaviour is considered multimodal if there is a temporal overlap between at least two different communicative elements (gestures, vocalizations, and/or gaze) at some point during the behaviour.

[50] The MacArthur Communicative Development Inventory (CDI) is a questionnaire used to assess language development in children. Parents fill it out, indicating the words and sentences their children understand and produce, as well as other aspects of communication and language comprehension. In this study, the CDI was

apparent in both gesture use and vocal behaviours. Children with TD exhibit proportionately greater use of reaching and conventional gestures[51] related to social routines than children with ASD. Conversely, children with ASD vocalize more than children with TD. Children with ASD often produce atypical vocalizations that may not be perceived as speech-like by adults. The study's analysis of communicative resources indicates that these children have a lower tendency to use combinations of two communicative elements, especially those involving eye contact. For instance, a child with typical development (TD) may point to an object and simultaneously look at an adult to attract their attention, using a combination of gesture, vocalization, and gaze. In contrast, a child with autism spectrum disorder (ASD) may only point to the object without integrating gaze into their communication. The study also found that the use of these three communicative elements increases with the development of vocabulary in children with TD. For instance, a child with typical development at an advanced stage of language may combine an object (gesture), a related word (vocalization), and eye contact with an adult (gaze) in a single communicative attempt. However, in children with ASD, the development of multimodal communication is not as apparent. They often struggle with tasks that involve integrating three or more elements, especially when it comes to gaze integration.

By comparing the results of recent studies, such as Murillo et al. (2021a), with earlier research, such as Meltzoff & Borton (1979), and further studies by Murillo (2021b), an important aspect of multimodal communication emerges: the capacity for coherent integration of information. The ability to integrate information from different sources in multimodal communication is important both perceptually and productively. Initially, in the 1980s, there was discussion of a general ability to integrate information, but more recently, the issue has also been related to the number of elements a subject can handle. Research on children with ASD and LD indicates that they retain the ability to integrate information in multimodal communication. However, their ability to combine elements becomes challenging when the number of elements to be integrated exceeds a certain threshold (as

---

used to match children with ASD to children with TD based on their language development, allowing for more balanced comparisons between the groups.

[51] Reaching gestures in children with TD refer to a type of communicative gesture in which the child extends his or her arm and opens his or her hand toward an object, indicating a desire to interact with or call attention to that object. This gesture is one of the basic ways that children begin to communicate their intentions and interact with the world around them before verbal language is fully developed.

observed in Murillo et al.'s 2021 study, where the threshold was 3). As previously mentioned, both the feature discussed and the multimodal shift feature are crucial in addressing the epistemological and metaphysical questions surrounding ToM. However, it is important to note that the research conducted has never directly addressed the relationship between multimodal communication and ToM. Nevertheless, if we consider multimodal shift as an indicator of ToM ability, the ability to integrate information coherently can be seen as a precursor to the development of both linguistic ability and ToM. With these foundations in place, we can now delve into these relationships. In the next chapter, I will build a theoretical framework capable of unraveling the complex issue of assessing ToM capacity in non-linguistic animals and understanding its underlying mechanisms.

3 Multimodal mind theory: a new perspective

The previous section attempted a reconstruction of the history of the study of animal language and communication since the ancient Greek period. The first approach was to identify lògos as the feature that distinguishes human mind and speech from other animal mind and vocalizations, emphasizing the human ability to articulate sounds. The study of vocalizations is divided into three parts: phoné, diàlexis, and psòphos. This tripartition was first proposed by Plato, Xenophon, and Hippocrates, and later revised by Aristotle. Additionally, the Stoics theorized the difference between lògos prophorikòs and logos ediàthethos, which further impacted the study of mind, language and communication of humans and other animals. These concepts continued to be influential throughout the Middle Ages. Only in modern times has interest in gestural language, in addition to vocalizations, begun to emerge, particularly in the field of human communication. Conrad Amman and Charles-Michel de l'Epèè are notable figures who contributed to the development of communicative techniques for the deaf. Regarding nonhuman animals, the distinction between the two concepts is made even sharper by Renè Descartes, who introduced the dichotomy between res extensa and res cogitans. This excludes animals from participating in mental life and, consequently, in language, which is understood as its direct manifestation. This approach has influenced several authors, most notably Chomsky, who reiterated in the 20th century that language and linguistic creativity are exclusive characteristics of the human species. This idea has resonated so strongly that funding for research on language and animal communication has diminished, if not ceased. However, Gassendi, Condillac, and Darwin had already recognized that animal communication occurs in forms other than vocalization alone, introducing rudimentary concepts of multimodality. Darwin emphasizes the use of vocal, gestural, and physical signals in animals. Therefore, the study of communication in animals has increasingly focused on multimodality in the 21st century. Thus, there has been a shift from a unimodal perspective centered primarily on vocalization, and only secondarily and more recently on gestures, to a multimodal approach that allows for a deeper and more integrated understanding of animals' communicative abilities and their relationship to cognitive abilities. Pioneering studies on multimodality are already being conducted by the founders of ethology. Studies of the bees' dancing, the supra-normal stimuli, and the way different species integrate different signals, are all examples of how the phenomenon of multimodality functions. As well as a review of the main scientific evidence, the debate surrounding the definition and classification of multimodal signals has been highlighted, showing how definitions depend on the conceptual framework used for classification and how various vagaries of terminology make the creation of a unified framework difficult. Despite the difficulties, it has been argued that by using frameworks that emphasize the importance of the perceiver's role in communication and shifting the focus from the signal transmitter to the receiver, it is possible to explore the presence of Theory of Mind (ToM) in animals that exhibit multimodal communication. This research focuses on two mechanisms related to multimodality: multimodal shift and cross-modal binding. The first mechanism can potentially indicate the presence of Theory of Mind (ToM) in non-linguistic animals. The second serves as a precursor to the development of linguistic and communication abilities as well as ToM in animals. This chapter will address these conceptual questions and attempt to construct a theoretical framework capable of addressing the well-known epistemological and metaphysical issues. Section 3.1 clarifies which multimodal shift executions may indicate the presence of ToM and why. Section 3.2 argues that the correlation between second-order linguistic and cognitive abilities is not causal, but rather co-developmental with the cross-modal binding ability. Section 3.3 proposes two mental experiments to support the possibility of a such theoretical framework. These experiments aim to clarify how the development of complex cognitive abilities cannot emerge in unimodal processing contexts or in the total absence of perceptual faculties. Section 3.4 demonstrates how current neuroscience knowledge is consistent with the proposed theoretical framework.

3.1 The multimodal shift as an indicator of Theory of Mind

The previous chapter discussed the characteristics and classification method of multimodal shift. It was concluded in section 2.4 that a particular type of multimodal shift could indicate the presence of ToM ability in animals. Multimodal shift can occur due to both the informant and the percipient understanding each other's perceptual or psychological state. If applicable, the multimodal shift would aim to either simplify or complicate the process of selective attention[52]. The process of selective attention, as proposed by Lev-Ari et al. (2022) within the ecological approach, falls within the behavioural spectrum available to both humans and other animals, following an evolutionary and developmental line through distinct stages. The process of selective attention can be simplified or complicated due to abiotic environmental factors. This can occur in situations where there is cooperation or competition among conspecifics, or during the implementation of anti-predatory strategies. Facilitating or complicating the process of selective attention in the cases mentioned requires the individual performing the multimodal shift to have ToM skills.

Not all instances of multimodal shift, however, serve the purpose of facilitating or complicating the process of selective attention or require a ToM capacity. To be considered a valid indicator of ToM, multimodal shift must exhibit specific characteristics. If it does not exhibit these characteristics, it cannot be considered a reliable indicator of ToM. To test for ToM, it is therefore necessary to draw a line to distinguish valid multimodal shifts from those that are not valid for indicating ToM. By discriminating valid multimodal shifts from those that are not, it will then be possible to design consistent experiments that can exploit these behaviours to indicate the presence of ToM in non-linguistic animals.

A crucial aspect of an experimental protocol employing the multimodal shift mechanism is to guarantee its applicability irrespective of the species-specific characteristics of non-linguistic animals. Multimodal shift is a widespread phenomenon in different animal species, but it has specific peculiarities in each species. It is based on the communicative repertoire available to different species. Therefore, the experimental protocol must be adapted to the species-specific needs of the animal being studied. Furthermore, if the experimental protocol were to be applied to various animal species within the same taxon, there may be compelling evidence to suggest the existence of ToM in all animals within that taxon that are capable of exhibiting a particular class of multimodal shift without being subjected to the experimental protocol. This is due to the shared neural structures and cognitive mechanisms within a taxon

---

[52] Selective attention is the process of prioritizing and maintaining attention on specific stimuli while inhibiting the processing of competing or irrelevant stimuli. This can be exemplified by the cocktail party effect, where in a crowded party, if someone says your name, you immediately bring your attention to that conversation, actively following it and ignoring the background noise. Deficits in selective attention skills have been associated with several clinical conditions, including autism spectrum disorders. This is particularly relevant when it comes to using and understanding multimodal signals. It is important to note that this association is not limited to autism spectrum disorders.

resulting from a common ancestry. If research indicates that a multimodal shift in some species of a taxon is related to ToM, it is reasonable to assume that similar neural structures and cognitive processes may support similar functions in other species of the same taxon. Additionally, within a taxon, species often undergo both convergent evolution (independently developing similar traits) and divergent evolution (variations in a common trait). If a mechanism such as multimodal shift is indicative of ToM in one species, it is possible that similar mechanisms could be present in other related species. This could reflect both convergent adaptations to complex social life and variations on a common evolutionary theme. Species within a taxon often share key aspects of social and communicative behaviour due to their phylogenetic inheritance. If a multimodal shift has been identified as a key component of social communication in some species and is associated with ToM, it could also be a reliable indicator in other species of the same taxon, given the likely presence of similar communicative and social systems. However, it is important to note that arguing that animals capable of exhibiting a specific multimodal shift are also endowed with some level of ToM does not necessarily imply that animals that are not capable of exhibiting it do not possess ToM. In their case, the presence of ToM will simply not be empirically testable with the experimental protocol that will be proposed. Let us review the characteristics of the ToM-indicative multimodal shifts.

When evaluating a ToM-indicative multimodal shift, the first consideration is whether multimodal shift is consistently exhibited in response to specific stimuli, regardless of context or the perceiver receiving the message. It is unclear whether multimodal shift occurs as an invariant response to stimuli or not. The data available do not provide a definitive answer to this question, which undoubtedly requires further investigation. However, to ensure objectivity, it is necessary to exclude from the characteristics of ToM-indicative multimodal shifts any multimodal shifts that is exhibited for mate-choice or courtship-related reasons. Furthermore, it is important to exclude those multimodal shifts that are exhibited widely throughout the population and not at the individual level (Partan et al. 2010; Hodl & Amézquita 2011; Belwood & Morris, 1987; Strauß & Lakes-Harlan, 2014). Mating behaviours are typically stereotypical and driven by genetic and instinctive mechanisms, with individual variation often limited to a species-specific range. Recognizable and rigid patterns suggest less influence of complex cognitive processes, such as ToM, which would imply greater flexibility and adaptability to the perspectives of others. An example of spiders' behaviour during courtship is illustrated in section 2.4 (Jackson, 1977, 1992; Taylor et al., 2005; Wilgers & Hebets, 2011; Scheffer et al., 1996; Gordon & Uetz, 2011). In low light conditions, spiders of a particular species, which typically use visual signals to communicate with conspecifics, switch to using vibratory signals. This shift in communication mode is an adjustment to the environmental context, but it still follows a recognizable pattern specific to the species and is likely influenced by genetic and instinctive mechanisms that are challenging to correlate with ToM activity. In the second case, if a behaviour such as multimodal shift occurs uniformly throughout a population, it is more likely to be the result of an invariant response to stimuli rather than a sophisticated understanding of the mental state of others. In contrast, if the behaviour is exhibited differently by individuals and is well adapted to the specific ecological context, this might indicate a greater likelihood of ToM, as it suggests an understanding and adaptation to the circumstances and needs of individuals. For example, squirrels in urban environments, as seen in Sections 2.3 and 2.4 (Partan et al. 2010), can be considered as exhibiting population-level behaviours. These animals rely more on visual signals than auditory ones in response to surrounding environmental noise. The multimodal shift is an adaptation to context, suggesting the presence of an invariant response to stimuli rather than a sophisticated understanding of another's mental state. This indicates that the animals' response is not influenced by subjective evaluations.

The second defining characteristic of a ToM-indicative multimodal shift is the nature of the signals that comprise the multimodal signal and their interrelationship. (i) Firstly, the components of the multimodal signal must be redundant. The informant must aim to convey certain information and, to facilitate the perceiver's understanding, change sensory channels while conveying the same message. Many birds use both auditory signals, such as calls or songs, and visual signals, such as wing or body movements, to attract attention (Cai & Dent, 2020; Akcay & Beecher, 2019). This combination of signals conveys the same message through different channels, increasing the likelihood of the message being received and understood by the perceiver. In section 2.3, we discussed a multimodal signal classification method based on the study of the meanings of transmitted signals. This method is considered one of the most problematic, calling into question Quine's (1992) principle of untranslatability. When referring to redundant signals, it is important to note that the interpretation of the signal is not relied upon, but rather the behavioural response of the percipient. If the percipient responds to different signals with the same behavioural response, the message conveyed by the different signals is the same. When stating that the message is the same, it is important to avoid interpreting or translating its meaning, as this can be problematic when classifying signals. Instead, we rely on the implicit assumption that if different signals in the same context influence the behaviour of the recipient in the same way, then the message they convey is the same. Although we cannot fully understand the nature of this message, we can still analyze its effects.

(ii) Secondly, redundant signals must be freely related to each other. Redundant signals should not be the necessary result of the production of any other component signals in the multimodal signal. An example of redundant signals freely related to each other can be seen in chimpanzees (Ross et al. 2014) during play, where a chimpanzee may emit a playful cry (acoustic sign) while jumping (visual sign). Although these signs may appear together, they are not necessarily connected. For example, a chimpanzee might jump without emitting a cry or vice versa. However, as pointed out in section 2.3, this distinction is not sufficient. Two signals that are freely related could still be the result of an invariant response to given stimuli. A clear and concise functional distinction should be made between signals that are intentional or unintentional. A signal can be considered unintentional if it is produced as a direct consequence of the emission of another signal or a given stimulus. Conversely, a signal can be considered intentional if it is produced independently of the emission of another signal and its emission cannot be explained as an invariant response to a given stimulus. In short, if the first condition is needed, that the multimodal shift does not occur in response to stimuli and furthermore the multimodal shift is based on redundant and free signals then signals can be defined as intentional. In Section 2.3, it was noted that the use of the concept of intentionality to distinguish between cues and signals was problematic. It is important to note that the use of the term 'intentionality' in this case is not problematic. We are referring to signals that have evolved to play a role in communication, even if we apply the contested distinction between cues and signals. The free relationship between signals suggests that the display of multimodal shift because of ToM activity may be found in possible attempts by an informer to deceive a perceiver. In such cases, the informer may not exhibit multimodal shift at equal disturbance in contexts where multimodal shift is exhibited if it can bring them some kind of advantage. Chimpanzees, for example, live in highly structured societies and have complex social rules, including those involving food sharing. In some chimpanzee groups, an individual who finds food is expected to communicate it to other members of the group, particularly dominant individuals, through clear and distinctive vocalizations. However, an individual wishing to keep food for themselves may decide not to make any vocalizations, an action that, if discovered, would imply punishment. Alternatively, it might be plausible for the chimpanzee to adopt more discrete signals, such as visual instead of auditory, to reduce the possibility of attracting the attention of the entire group. The use of such signals, rather than their absolute absence, could provide the individual with an alibi if discovered.

The third indicator of ToM is a multimodal shift that is performed by an informer only in the presence of potential perceivers, whether they are conspecifics or belong to other species. It is not necessary for the perceiver to be perceived at the time the multimodal shift is performed. What is important is that the informer is aware of the presence of a perceiver, even if it cannot be inferred directly from its surroundings. It is expected that when an environmental condition rapidly changes and reaches a certain threshold of disturbance on a sensory channel, such as the visual channel, the informer will perform a multimodal shift to communicate with the perceiver, even if the informer can no longer observe the perceiver. In this case, the informer was able to deduce the presence of the perceiver without directly observing the perceiver, as well as what the perceiver was able to perceive about their surroundings. The observer then adapted their communication accordingly. A practical example of this behaviour can be seen in nesting birds. If a bird needs to communicate with its mate or chicks and the visual channel is obstructed, it may switch to acoustic signals to maintain communication. This adaptation also occurs when the bird cannot see its mate or chicks directly but knows they are in the area and need to communicate. As a corollary, a multimodal shift does not have to exclusively involve the communication aspect related to perception by the perceiver to be considered an indicator of ToM. In other words, it does not have to be a perceptual multimodal shift. Focusing on one feature of a signal rather than another, in certain contexts, cannot be considered an indicator of the presence of ToM. For a multimodal shift to be an indicator of ToM, it must involve both the perceptual and productive aspects. For instance, squirrels in urban environments adapt to the noisy surroundings of cities by altering their communication methods (Partan et al., 2010). In urban environments with high noise density, squirrels rely on visual signals rather than auditory signals to communicate. This is because background noise can easily overpower auditory signals, such as calls. However, it is important to note that squirrels do not necessarily consider how their signals will be perceived by specific conspecifics. It is not the speakers, but rather the listeners who rely on the visual aspect of communication when they are unable to clearly hear the vocal message. This reduces uncertainty in the message but does not involve an understanding of the speaker's perspective or intentional modulation of communication.

To summarize, identifying multimodal shift as an indicator of ToM in non-linguistic animals requires a multifactorial analysis of communicative behaviour. Initially, it is crucial to determine whether the multimodal shift represents an invariant response to specific stimuli. If an animal systematically changes its mode of communication in response to specific stimuli, without regard to the context or recipient of the message, such a change may not be indicative of ToM. A crucial aspect of the multimodal shift is the redundancy and independence of the signals involved. Intentional and unintentional signals should be functionally distinguished, with the former being produced independently of other signals and not as a direct response to a particular stimulus. The third important feature pertains to the display of multimodal shift by the informer only in the presence of potential perceivers. It is crucial that the informer is aware of the presence of the perceiver, even if it cannot be inferred directly from the surroundings. This implies that the informer must be able to infer both the presence of the perceiver without direct observation and what the perceiver is able to perceive of the surrounding environment. Finally, it is essential that a shift towards multimodality not only involves the perceptual aspect but also the productive aspect of communication. A multimodal shift that is purely an adaptive reaction to the environment, without active modulation and understanding of others' perspectives, cannot be considered an indicator of ToM.

Briefly, if an animal’s (A’s) multimodal shift X is (i) not an invariant response to a stimulus, (ii) is relying on redundant and free signals (iii) is performed only in presence of a perceiver, and (iv) involves both the production aspect and the perceptual aspect of communication, then A has ToM.

Based on the data provided by Partan (2017) regarding the species capable of exhibiting multimodal shift (section 2.4), it appears that out of the sixteen species[53], only one exhibits all

---

[53] The sixteen species in which multimodal shift has been detected are, jumping spiders (*Phidippus johnsoni*), wolf spiders (*Schizocosa ocreata* and *Rabidosa rabida*), Alpine newt (*Mesotriton alpestris*), Stickleback fish (*Gasterosteus aculeatus*), Humpback whale (*Megaptera novaeangliae*), Eastern grey squirrel (*Sciuris carolinensis*), Hylodinae (*Hylodes*), Myobatrachidae (*Taudactylus*), Hylidae (*Litoria*), Ranidae (*Staurois*),

the necessary features for ToM detection through multimodal communication. Two other species exhibit all the characteristics mentioned, except for the components of the multimodal signal, whose redundancy is unknown. Assuming their classification as redundant, these two species, unlike the first one, will be considered only as likely candidates, and not as ideals, of the presence of ToM. The first case refers to the chimpanzee (*Pan Troglodytes*) (Tomasello et al. 1994; Leavens et al. 2010). In the second case, the species mentioned are the humpback whale (*Megaptera novaeangliae*) (Dunlop et al., 2010) and the California ground squirrel (*Otospermophilus beecheyi*) (Rundus et al., 2007). It is worth noting that chimpanzees exhibit multimodal shift not only in communicative contexts with conspecifics but also in interspecies contexts when interacting with their human keepers. The exhibition of multimodal shifts in interspecies contexts suggests that ToM can be used to explain the psychological cause of other creatures' behaviour, regardless of whether they belong to the same species. The use of ToM to explain the behaviour of other living things also implies that our tendency to anthropomorphize other animals may be a shared trait with other species that possess ToM. Just as humans anthropomorphize the behaviour of other animals, we might say that chimpanzees pitecomorphize[54]. The fact that other species exhibit multimodal shifts in interspecies contexts, such as in the case of some antipredatory strategies, raises the question of whether interspecificity might be another requirement for a multimodal shift to be interpreted as an indicator of ToM. However, it is more appropriate to exclude it. The reason for this issue is related to an individual's ability to recognize abilities in others beyond their own. Returning to the chimpanzee example, the chimpanzee could exhibit an interspecies multimodal shift because it hypothetically has a ToM, but it may not have the ability to detect its absence in individuals other than its own conspecifics. This is the concept referred to as anthropomorphization. Many people tend to believe that animals possess ToM as a matter of naive psychology, even though the presence of ToM in animals other than humans has not yet been demonstrated. On the other hand, it is possible that both humans and animals can recognize abilities in other animals that are different from their own. If this were the case, it would be necessary to determine whether the interspecies multimodal

---

Chimpanzee (*Pan troglodytes*), California ground squirrel (*Spermophilus beecheyi*), Chickens (*Gallus gallus*), Katydids (*Tettigoniidae*) and Cicada (*Cicadidae*).

[54] Since 'anthropomorphize' comes from the Greek words 'anthropos' (man) and 'morphē' (form) and refers to the attribution of human characteristics to other beings or objects, the term 'pitecomorphize' could be created using the Greek word 'pithekos,' which means 'ape.'

shift is not occurring due to a lack of ToM in the individual or because the individual does not attribute ToM to the other species it is interacting with. The issue of determining whether interspecies multimodal shift occurs through the attribution of ToM to other species can only be resolved once the methodology for attributing ToM to a non-linguistic animal has been established. Consequently, interspecies multimodal shift cannot be included among the variables employed in the construction of a protocol suitable for testing ToM in non-linguistic animals.

After establishing the necessary characteristics for a multimodal shift to indicate the presence of ToM, the question arises of how to utilize ToM-indicative multimodal shift to solve the epistemological problem. As previously mentioned, current tests for determining ToM in non-linguistic animals have failed to overcome the issue of double interpretability of results. Regardless of the experiment's outcome, it can be interpreted as evidence of ToM, or as a simple reading of the behaviour of others, without making any inferences about the minds of others. Although there are currently no tests that can overcome the issue of dual interpretability of results, it is also true that no test has been designed to utilize multimodal shift as a crucial component of the experiment. However, it is possible to design experimental protocols that can overcome the empirical problem and prevent double interpretation of the results by leveraging multimodal shift and adhering to specific conditions. The experimental protocol I propose is closely related to the Appearance-Reality Mindreading (ARM) theory developed by philosopher Robert Lurz (2011), as well as the specific protocols he devised based on that theory. The following chapter will discuss the importance of the ARM theory and these protocols at length. The explanation of the experimental protocol using the concept of multimodal shift will be deferred to Chapter 4.

In the previous chapter, it was hypothesized that multimodal shift could be an indicator of the presence of ToM in non-linguistic animals, and that the ability to integrate a certain amount of information coherently could be considered both a precursor and an indicator of the development of both ToM and the communicative and linguistic abilities of living beings. The following section will expand on this hypothesis to establish a theoretical foundation for the experimental protocol outlined in the subsequent chapter.

3.2 Cross-modal binding as a precursor and indicator of ToM, language and multimodal communication development

The studies reviewed in section 2.5 on multimodality in infants and neurotypical subjects indicate that a crucial aspect of multimodal communication is the capacity to integrate signals coherently, both perceptually and productively. This integration of different signals is referred to as cross-modal binding. Although not discussed previously, it is evident that the ability of cross-modal binding precedes the development of multimodal communication and, consequently, multimodal shift. Multimodal communication relies on the integration of multiple sensory channels, such as visual, auditory, tactile and so on to influence a percipient's behaviour. Multimodal communication requires the ability to process and coordinate signals from different sensory channels. Cross-modal binding is necessary for this type of communication, and it is not limited to humans. Several studies, as noted in the previous chapter, demonstrate that nonhuman animals use multimodality to communicate. Therefore, they must possess cross-modal binding ability. Cross-modal binding is essential for consistent perception of a certain state of affairs and for adapting communication to the ecological context. When an organism perceives a multimodal signal, such as the sight and sound of another animal approaching, the different sensory signals are processed by the respective sense organs and transmitted to the brain. These sensory inputs are processed separately and simultaneously, and then integrated into a cohesive perception that reflects the reality of the observed event. This integration enables the perception of an event as a unified entity, despite the presence of multiple sensory inputs.

Research suggests that nonhuman animals, including mammals, can integrate perceptions into complex cognitive representations (Regolin & Loconsole, 2023; Jaros & Pudil, 2020). Thus, some nonhuman animals are capable of tracking correlations between multisensory information, rather than simply combining it at the moment of perception. For instance, dogs are a prime example. A dog integrates both sight and smell to identify its owner by keeping track of the correlation between the two senses. By associating the visual image of the owner with their characteristic smell, the dog forms a representation that allows it to recognize the owner even in the absence of a direct visual stimulus (Andrews et al. 2022). Representations are mental images or concepts perceived by our minds[55]. Some nonhuman animals can correlate signals that co-occur in time and space and memorize those co-occurrences. In other words, the tendency to form links between co-occurring stimuli helps animals quickly and efficiently understand and act on their surroundings. The way animals link sensory signals that co-occur in time and space can be interpreted as the application of heuristic logic. Heuristics are cognitive strategies that simplify decision-making, allowing individuals to draw quick and often accurate conclusions with minimal cognitive effort. They are particularly useful in situations that require quick responses or when limited information is available. Heuristics are based on the recognition of patterns or correlations that have occurred frequently in the past. For instance, an animal can learn to associate certain sounds with specific visual sources, such as a particular bird's song that corresponds with its visual presence. This association allows the animal to infer the presence of another individual, such as another type of bird, when it hears a new sound similar to the bird's song. Nonhuman animals, like humans, do not need to process and analyze every aspect of sensory information in detail. They rely on pre-existing connections and associations, formed through past experiences, to make efficient decisions.

As mentioned earlier, cross-modal binding plays a crucial role in multimodal communication and language acquisition, especially in humans, including infants and adults (Marinis, 2018; Cuskley & Kirby, 2013). Cross-modal binding aids in word learning by enabling the formation of associations between unfamiliar words and their semantic referents (Wang et al, 2017). Semantic referents are the real-world entities to which a word or phrase refers[56]. In the context of child language development, cross-modal binding enables children to associate heard words (sound signals) with specific objects, actions, or concepts (visual signals or representations). For instance, when a child hears the word 'ball' and simultaneously sees a ball, cross-modal binding helps them connect the sound of the word ('ball') with the semantic referent (the actual ball), facilitating vocabulary acquisition. Cross-modal binding has been

---

[55] The representational theory of mind is a philosophical theory that posits that the mind perceives only mental images or representations of external material objects, not the objects themselves. Representations play a crucial role in ToM, according to the representational theory of mind. are formed in our minds when we attribute mental states to others. These representations guide our interactions with others, enabling us to predict or explain their actions. ToM in this sense is highly dependent on our ability to form and interpret representations.

[56] For example, in the sentence "The cat is on the table," "the cat" and "the table" are the semantic referents because they refer to real entities in the world.

observed in second-language word learning in adults, indicating that it plays a similar role in learning new words for both familiar and unfamiliar semantic referents. For familiar referents, adults only need to associate a new verbal label, while for unfamiliar referents, they need to form a new link between a new word and a new concept or object. For instance, consider an adult who is learning a second language. In this context, the effective application of cross-modal binding is necessary to associate heard words, such as new terms in a foreign language, with familiar concepts or objects. For instance, when an adult learns the Spanish word 'perro' (dog), they link it not only to the sound but also to the mental representation or visual perception of a dog. Cross-modal binding enables the learning of new words and the association of new terms with known concepts. The literature suggests that there are correlations between lower cross-modal binding ability and lower language development, highlighting the key role that cross-modal binding plays in language acquisition. For instance, individuals with dyslexia exhibit decreased cross-modal binding, which can be diagnosed through difficulties in learning and word recognition (Jones et al., 2013). Therefore, cross-modal binding is crucial not only for language acquisition but also for its maintenance and development over time. It is important to note that studies on the relationship between cross-modal binding and language, as discussed in section 1.4, may be biased as they often only consider language from a speech and therefore unimodal perspective. It is worth considering sign language, which varies between languages such as Italian Sign Language and English Sign Language. It is clear that cross-modal binding plays a role in allowing us to associate signs with concepts or semantic referents when learning a new language. This process is not limited to the verbal context but also occurs in the broader multimodal context. Another important point to emphasize is the inference of similar cognitive processes in non-human animals, although there is less literature available on this topic. This inference, that comparable cognitive functions exist in nonhuman animals, is especially plausible in species that are renowned for their complex communication systems.

In addition to recognizing the fundamental role that cross-modal binding plays in the development of the communicative and linguistic faculties, it is worth considering whether cross-modal binding also plays a role in the development of ToM skills. Both language and multimodal communication, as well as ToM, are cognitive processes that involve the integration of multiple sources of information. Therefore, it is possible that all of these faculties are related to cross-modal binding ability. To demonstrate that ToM requires the use of multiple information, a thought experiment is proposed in Section 3.3. For now, let us accept this idea as true. If cross-modal binding does indeed play a role in the development of ToM, then we should observe correlations between lower cross-modal binding ability and lower ToM ability. However, no such studies have been found in the literature. As mentioned in Chapter 1, ToM cannot be demonstrated in non-human animals. Therefore, the studies mentioned in Section 2.5 on prelinguistic infants and children with autism spectrum disorder (ASD) and delayed language development (LD) are useful in attempting to correlate ToM development with cross-modal binding. These studies highlight two things: (i) the development of cross-modal binding precedes that of ToM. Meltzoff and Borton's (1979) study found that the ability of cross-modal binding was detectable in children as young as 29 days old, while ToM was detectable as early as 18 months of age (Meltzoff, 1999; Luo & Baillargeon, 2005, 2007; Onishi & Baillargeon, 2005; Onishi et al., 2007). (ii) The development of cross-modal binding may proceed at different rates in different individuals and may even stop. As seen in section 2.5 and according to a study by Murillo et al. (2021a), individuals with ASD and LD have difficulty integrating more than three signals simultaneously. Considering that ToM may depend on the development of cross-modal binding, it is reasonable to expect that children with ASD and LD would have a lower degree of ToM development compared to children with typical development (TD). An initial review of several texts in the field of psychology indicates that individuals with ASD lack ToM as an established fact (Coon, Mitterer, & Martini, 2018; Kellogg, 2007; Kirk, Gallagher, Coleman, & Anastasiow, 2008; Mash & Wolfe, 2015; Myers, 2009, 2012; Sigelman & Rider, 2017). According to a 2019 review by Gernsbacher and Yergeau, over 75% of the top 500 articles on Google Scholar related to the relationship between ToM and ASD tend to argue that people with ASD lack ToM. However, the assertion that individuals with ASD lack ToM present a challenge when attempting to substantiate the association between cross-modal binding and ToM. Less development of cross-modal binding should result in less development of ToM, rather than a complete absence of ToM.

This is where Gernsbacher and Yergeau's review becomes more relevant. The researchers argue that the available empirical evidence often does not support the assumption that individuals with ASD lack ToM, despite this assumption being widely accepted. This assumption is also used by several psychologists in state and federal court cases in the United States (Carter v. Superintendent 2011; New Jersey v. Burr 2007; United States v.

Geanakos 2017). Research cited in Google Scholar suggests that individuals with ASD may lack ToM, often citing previous research by Baron-Cohen (1988, 1989a, 1989b, 1989c, 1995). However, Gernsbacher and Yergau point out that a substantial amount of data, collected by researchers not primarily focused on ToM and ASD, contradicts this claim. Difficulties with ToM tasks, for example, are not unique to ASD; children with conditions such as specific language impairment and Down syndrome face similar challenges, casting doubt on the notion that ToM deficits are unique to ASD. Additionally, it is inaccurate to claim that all individuals with ASD lack ToM skills. Some studies (Bailey et al., 1996; Bauminger & Kasari, 1999) have found that some individuals with ASD can successfully complete basic ToM tasks, suggesting that ToM deficits are not universal in ASD. However, the reproducibility of basic ASD and ToM studies has been problematic. Attempts to replicate the original results of studies conducted by Baron-Cohen et al. (1986) often fail, casting doubt on their reliability. Furthermore, the instruments utilized to measure ToM often produce inconsistent results, suggesting that they lack convergent validity. An individual's performance on a ToM task may not correspond with their performance on other ToM evaluations (Ahmed & Miller, 2011; Lukito et al., 2017). Lastly, the practical significance of ToM tasks is currently being questioned. The predictive ability of social-emotional functioning, autistic traits, empathy, emotional understanding, or everyday social skills in individuals with or without ASD has not been convincingly demonstrated (Kunihira et al., 2006; Ronald et al., 2006).

Gernsbacher and Yergeau's review (2019) is a crucial contribution to the ongoing discussions about ToM and ASD. Their insights reveal that the absence of ToM is not universal among individuals with ASD. This revelation, combined with challenges in replicating core studies, inconsistencies in ToM assessment tasks, and the questionable predictive power of these tasks, suggests that ToM might manifest differently or be less developed in individuals with ASD, rather than being completely absent. This hypothesis proposes a potential link between cross-modal binding abilities and ToM. The suggestion is that the development of cross-modal binding may influence ToM abilities. Investigating this relationship could provide insight into the social and cognitive skill development of children with ASD. The hypothesis proposes that limitations in cross-modal binding may not eliminate ToM entirely but could lead to its reduced or altered expression. The observed variability in ToM skills among children with ASD may be explained by this nuanced view, which is more consistent with empirical findings. It is important to note that even if ToM is not readily observed in individuals with ASD, it does not necessarily imply a total absence of such cognitive skills. Rather, individuals with ASD may exhibit different levels of ToM ability that are not identifiable by the tests used to verify its presence. In studies related to Theory of Mind, particularly those involving animals or individuals with nonstandard communicative-linguistic abilities, the primary issue is often epistemological rather than inherent in the cognitive abilities of the subjects being tested. Regarding non-human animals, if we acknowledge that ToM is not necessarily absent, but can manifest itself in different ways depending on the level of development of cross-modal binding, then this gradualness in the manifestation of ToM can be found not only in humans, but also in non-human animals. This is likely to be true in non-human animals, at least in those closest phylogenetically.

If both multimodal communication, language and ToM all depend on the development of cross-modal binding, it is likely that the long-standing debate about the relationship between ToM and language will shift.

As mentionend earlier in Chapter 1, there are several schools of thought on the relationship between language and ToM (de Villiers, 2021). (i) The conduit view suggests language merely expresses pre-existing concepts formed without the involvement of language. (ii) The cultural view posits that language transmits cultural understandings of theory of mind, and that children develop this understanding through discourse. (iii) Language as a cognitive tool theory suggests language delays can hinder theory of mind development. (iv) Lastly, the representational view suggests language structures aid in representing complex events and reasoning, with complement structures being strong predictors of false belief understanding in children. The aforementioned theories are further distinguished by the school of thought that ToM development precedes language. In this case, the development of language is contingent upon the development of ToM. Each of these schools of thought, except for the first one, posits a correlation between language development and ToM development. However, none is able to provide a definitive explanation of the causal link between the two. While each school of thought offers evidence to support its position, there is also evidence that challenges the plausibility of its claims.

Instead, the argument that both ToM and language depend on the development of cross-modal binding is a compelling one. It succeeds in theorizing clearly because a correlation is identified between the degree of language development and the degree of ToM development.

The hypothesis that there is a correlation between cross-modal binding, language, and ToM becomes even more plausible as it can resolve controversies about contradictory data that arise in the analysis of empirical results. This is especially true when one wants to assume one of the other two main views, either that ToM depends on language or that it is a necessary prerequisite for language. For simplicity, let us refer to the view that ToM and language development depend on the level of development of cross-modal binding as a multimodal mind theory. For example, the multimodal mind theory succeeds in explaining the observed variations in ToM and language development in ASD and LD cases, as already demonstrated, among the studies already mentioned, in contrast to the aforementioned schools of thought as will be elucidated a little further on. Chapter 5 discusses cases that the schools of thought on the interdependence of ToM and language cannot explain, but which can be explained by the multimodal theory of mind. According to this theory, ToM is not dependent on language, but rather relies on the integration of various sensory information. Other animals, like humans, process multisensory information to understand and explore their ecological context. This suggests that they can develop a form of ToM without the need for language because they have a cross-modal binding capacity. The degree of ToM development is greater in subjects who can process and integrate information from different sensory modalities. Cross-modal binding can be understood as bipartite. (i) The capacity to process multiple signals simultaneously. (ii) The ability to recognize patterns among co-occurring signals through memory and experience. Both factors contribute to the overall level of cross-modal binding and subsequent language and ToM development.

According to the multimodal mind theory, language and other communication systems may serve the purpose of rendering a potentially unconscious phenomena (like ToM) accessible to conscious investigation. Unconscious phenomena do not exist because of language, but rather can be investigated through language. This point, although not central to the purpose of providing a theoretical basis for multimodal shift as an indicator of ToM in non-linguistic animals, will come in handy in Chapter 5. Specifically, if we refer to the role that language plays according to the representational view, as proposed by De Villiers (2022), it is often argued that to have ToM, it is necessary to be able to think about mental states consciously. Furthermore, according to the representational view, to be able to think about mental states consciously, it is necessary to have language. The objective of this section is to lay the groundwork for an argument that multimodal mind theory has the potential to offer an alternative way to the view that ToM must necessarily be a conscious process that requires language. In order to support the hypothesis that ToM can be an unconscious process, however, it is first necessary to establish that cross-modal binding, on which ToM depends, can also be an unconscious process.

According to current theories of consciousness, cross-modal binding cannot occur unconsciously. To explain why, the concept of Global Workspace Theory (GWT) must be introduced. GWT, introduced by Baars in 1988, is one of the most influential theories of consciousness. The central idea that consciousness serves as a mechanism for providing global access, enabling the integration and sharing of information between otherwise independently operating functions, such as the specialized processing required by the different senses, has had a major impact in cognitive science research. The GWT has been widely accepted by many active researchers and theorists in the field. Several theories attempt to reconcile experimental data with models of cognitive and neurophysiological architectures. These theories differ considerably, but they all rest on a common foundation: the conscious access hypothesis (Baars, 2002). The conscious access hypothesis is present in many theories and is more explicitly articulated in GWT. GWT proposes that the brain is composed of a network of specialized processors that support sensory functions and motor control, among others. These processors operate largely independently and unconsciously. On the other hand, the global workspace is suggested to be widely distributed throughout the brain, mainly through cortical regions. It provides a mechanism through which information can be transmitted to different functional areas. Specialized processes compete for access to this workspace, allowing for the global distribution of information. The information we are conscious of are in the global workspace and subsequently are globally available. When multiple pieces of information are present in the global workspace, they can be integrated with each other (cross-modal binding). This is why, according to GWT, cross-modal binding cannot occur unconsciously. Without consciousness, the information from different sensory inputs is all disconnected from each other.

After explaining the conscious access hypothesis, it is worth noting that some recent studies challenge it by supporting the possibility of unconscious cross-modal binding (Shanks, 2010; Bushara et al., 2003). The 2018 study by Scott et al. builds on several previous studies (Pessiglione et al., 2008; Raio et al., 2012; Seitz et al., 2009; Duss et al., 2011; Henke et al., 2013; Tachibana & Noguchi, 2015) to demonstrate the acquisition of novel cross-modal associations between stimuli that are not consciously perceived. The study shows the capacity for unconscious associative learning in the auditory modality using a linguistic framework. The researchers then replicated the experiment for the visual modality. They subsequently adapted the experimental design to test the ability of unconscious cross-modal binding. After demonstrating this ability, they compared the results with those obtained from another test designed to determine if study participants responded differently by consciously processing information from the two different sensory inputs. The study found that conscious processing leads to faster processing but also a higher margin of error in classification compared to unconscious processing. Additionally, a final experiment was conducted without linguistic frameworks, which still showed cross-modal binding occurring at both conscious and unconscious levels. In summary, recent findings indicate that the unconscious plays a more complex and central role in processing sensory information, which challenges traditional theories that place consciousness at the center of sensory integration. If cross-modal binding can take place at the unconscious level, then there should be no theoretical limits to argue that ToM, if dependent on cross-modal binding, can also be an unconscious cognitive process.

To summarize, multimodal mind theory suggests that language and ToM are not mutually dependent. Instead, both language and ToM appear to depend largely on the degree to which cross-modal binding is developed. Multimodal mind theory indicates that the ability to integrate information from different sensory inputs is critical in determining the level of development of both ToM and language skills. In this analysis, the role of language is viewed differently. It is not just a means of communication or a phenomenon related to consciousness, but also a tool for bringing unconscious processes to the conscious level. Language serves as an interface that enables the exploration and understanding of phenomena that would otherwise remain hidden in the unconscious.

However, it is important to note that multimodal mind theory has some limitations and potential weaknesses. Firstly, the relationship between cross-modal binding and ToM and language development is neither fully understood nor supported by a unanimous consensus in the scientific community. Therefore, there is a risk of oversimplifying cognitive processes that are extremely complex and still the subject of intense research. Secondly, although the concept of cross-modal binding provides a reasonable explanation for the interaction between various sensory and cognitive systems, accurately measuring and identifying this phenomenon remains a challenge. How can we precisely quantify the extent to which cross-modal binding develops in different individuals? What are the objective parameters for evaluating this ability? In light of these difficulties and limitations, and in the impossibility of addressing them within this research project, I want to try to support what has been argued so far with the use of two mental experiments, subsequently analyzing what evidence from the field of neuroscience should support the different key points. Thought experiments can support the theoretical framework and clarify the usefulness of testing the presence of ToM in non-linguistic animals based on multimodal experiments to reduce the problem of dual interpretability of data.

3.3 Thought experiments of non-linguistic animal without sensory perception

The first thought experiment closely follows the structure of another well-known thought experiment in the history of philosophy: the flying man thought experiment. This experiment was proposed by the Persian philosopher Avicenna to argue for the independence of the mind from the body within the debate on mind-body dualism. In brief, Avicenna suggests imagining a man who is suddenly created without any history, memory, or experience. The term 'flying man' refers to a person who is suspended in the air without any physical sensation. According to Avicenna, even in this state of sensory deprivation, the person would still be aware of their own existence, demonstrating that consciousness is not dependent on the body or physical sensations. Upon revisiting Avicenna's thought experiment, it is proposed that cross-modal binding can occur unconsciously, and that the state of consciousness is subordinate to perceptual abilities and cross-modal binding. Not only consciousness, but also language and alternative communication systems depend on the same perceptual abilities and cross-modal binding.

Imagine an animal born without any internal or external sensory perception. It has no active senses such as sight, hearing, touch, taste or smell, nor can it perceive internal feelings or sensations. This animal, which we will call Alex for convenience, cannot have thoughts since there is no sensory input from which thought can emerge or on which concepts of any kind can be formed. Indeed, without perceiving stimuli such as thirst or hunger, Alex would be a mere biological entity, living in a vegetative state and dependent on external assistance for survival. Alex's existence is reduced to basic biological functions without the involvement of an active thought process. In the absence of interactions and experiences, and in a context devoid of thought, Alex would not have the means to develop a personality or identity. Personality and identity arise from the interaction between sensory experiences and individual responses to those experiences. Without the ability to distinguish one's body from the external world or to categorize the external world, one would not have a sense of self or personality. In the case of Alex, the lack of perception and thought makes it impossible to create communicative systems. A communicative system is a system of symbols that combine with each other following specific rules. This system can be based on articulated language, gestures, images, or mind maps. The ability to associate meanings with signifiers is necessary to use this system effectively. This means being able to associate a semantic referent with a counterpart that can represent it. This association assumes the presence of a type of thinking that enables the formation and interrelation of concepts. Symbolic systems are necessary for Alex to develop an internal language or a way of representing the world, which are considered essential for the experience of consciousness. Without them, Alex remains in a state of non-consciousness, unable to experience or reflect on the reality around him. However, what if Alex gradually acquires sensory abilities?

Imagine Alex gains the ability to hear[57]. He perceives sounds from the outside world but cannot yet produce them. However, at this stage, he cannot distinguish whether sounds come from himself or his surroundings. Without a feedback mechanism, such as sound production, Alex may not be able to understand the origin of sounds, leaving the distinction between himself and the outside world ambiguous. The next step in his development is the acquisition of the ability to vocalize, which is simply the ability to produce sounds. This change is crucial

---

[57] Neurobiologists from HSE University and the RAS Institute of Higher Nervous Activity and Neurophysiology have demonstrated that the human brain can unconsciously distinguish between even very similar sound signals during passive listening (Liaukovich et al., 2022). Although we might not always recognize these differences consciously, our auditory system is capable of detecting sounds at an implicit level. In their study, participants listened to sounds while researchers measured their brain responses using electroencephalography (EEG). The sounds were so similar that participants could explicitly distinguish them with only 40% accuracy. The brain responds differently to local irregularities (which can be detected without explicit attention) and global irregularities (which demand concentration and reflect a higher level of consciousness). Even when the sounds were barely distinguishable, the brain still exhibited these responses. Similar unconscious processes are also there for vision (Goodale & Milner, 2013)

because it provides Alex with a direct means of influencing and interacting with his sound environment. As Alex begins to produce sounds, he can experience the cause-and-effect relationship between his actions (vocalizing) and the resulting auditory perceptions. This helps Alex understand that some sounds are under his direct control while others are not. The ability to distinguish between self-produced sounds and external sounds is critical to the development of Alex's sense of self. This distinction allows him to begin to understand the separation between his being and the external world. With the simultaneous acquisition of hearing and speech, Alex can begin to learn from his interaction with the environment. This learning process is bidirectional. Alex not only receives information from the outside world but also begins to actively influence it. Although Alex has acquired hearing and the ability to produce sounds, he has not yet developed speech. However, he is able to categorize or signify sounds in a primordial form. This categorization is basic and relies on the direct association between sounds and actions or reactions. For instance, Alex can differentiate sounds based on characteristics like intensity or duration, but without assigning complex meanings to them. At this stage, Alex's thinking remains relatively simple and concrete. His cognitive abilities are limited to more direct and less abstract forms of thinking. Alex is capable of recognizing patterns and making simple connections. However, he struggles with developing more complex or abstract thinking due to his limited and poorly characterized understanding of concepts.

Now, imagine that Alex begins to develop vision, making it possible for him to perceive objects and events around him and develop a sense of space. With this new ability, he can locate the source of sounds he hears in space. For instance, he can visually associate the sound of a waterfall with the waterfall itself and link the song of a bird to the sight of the bird. Alex's ability to attribute sounds to specific sources improves his understanding of the environment. This enables him to recognize and differentiate objects and living things based on their auditory and visual characteristics. The attribution process occurs unconsciously, where visual and auditory information merge to form a unified and coherent perception of reality. The co-occurrence of information from different sense organs allows Alex to construct more complex categories. For instance, Alex can begin to group objects and sounds based on common characteristics, such as the types of objects that make certain sounds. This categorization and signification process is more advanced than at the stage when Alex had only hearing because categories can now be formed on a multisensory basis, integrating different forms of information. Additionally, Alex learns to see himself moving in space and to have a different perception of his body. The concept of self is broadened.

As Alex acquires different sensory perceptions, he continues the process of multisensory integration that occurs unconsciously. This integration allows Alex to connect information from different senses, creating a richer and more detailed understanding of the world around him. For instance, the association between a sound and its visual source, such as a bird's song and the bird itself, occurs automatically without conscious effort. These associations serve as the foundation for more intricate categories and a more cohesive perception of the external world. The development of language in Alex is heavily reliant on this multisensory integration. Language, in its broadest sense, is founded on the capacity to connect symbols (words) to particular concepts, objects, or actions, and the ability to relate them to one another by following specific rules (grammar). The integration of sensory information on an unconscious level provides the foundation for the creation of symbols and concepts. The ability to associate sounds with objects and perceive relationships between different sensory stimuli is crucial for the formation of language. In the same way, Alex's consciousness, as a subjective and reflective experience, emerges and develops from the integration of sensory information. Consciousness necessitates an understanding of oneself in the world, rooted in the integration of various sensory inputs. The ability to experience the world coherently and unified is a prerequisite for self-awareness and reflection on oneself and the environment. As one transitions from a state of nonperception to multisensory perception, it becomes clear how the unconscious integration of perceptions is fundamental to constructing an intelligible reality and developing higher cognitive abilities[58].

---

[58] Several real cases seem to support the thesis that the development of high cognitive abilities depends on multisensory integration. One such case is Samuel Gridley Howe, a U.S. physician and educator. Samuel Gridley Howe developed a groundbreaking educational approach while working with Laura Bridgman, a young woman who lost her sight and hearing due to scarlet fever. His methodology emphasized the importance of sensory integration and anticipated the principles of cross-modal binding. Howe used raised labels attached to objects, which allowed Laura to associate the tactile sensation of the letters with the object itself, establishing a basic form of touch-based language. The initial stage was pivotal as it provided Laura with a tangible means to comprehend and engage with her surroundings, despite her sensory constraints. Howe subsequently built upon this foundation by introducing the manual alphabet, a communication method that employs hand gestures to represent letters. This alphabet enabled Laura to articulate herself in a more refined manner, thereby facilitating more intricate communication. The interaction between Howe and Laura was important. Howe spent hours working with her, guiding her through tactile exercises and encouraging her to explore her environment. This process is similar to Alex's adaptation of her sensory abilities in the context of the thought experiment. Howe's strategy of providing immediate feedback and positive reinforcement whenever Laura correctly identified an

The second proposed thought experiment aims to support the thesis that multimodal communication and cross-modal binding enable and influence the development of ToM. In this experiment, we imagine two individuals, Bob and Charlie, living in a world where the only way to communicate is through a unimodal and unicomponent verbal channel, without any variation in tone or intensity, and without the possibility of using any non-verbal expression. Imagine a conversation between Bob and Charlie. Charlie informs Bob that he is happy. In a typical communication context, Bob would have access to various verbal and nonverbal signals - a smile, a cheerful tone of voice, an open body posture - to assess whether Charlie is truly happy or whether he is masking his true emotional state. However, in a world characterized by unicomponent communication, Bob must accept Charlie's statement at face value without the possibility of seeking confirmation or contradiction through other communication channels. ToM, as we know, is based on the ability to attribute mental and intentional states to others, often by analyzing inconsistencies or coherence between different forms of expression. In a typical conversation, discrepancies between verbal and nonverbal communication can guide us in understanding a person's true feelings or thoughts. However, in Bob and Charlie's situation, these nuances are inaccessible. If Charlie expresses sadness without any accompanying tonal variation or facial expression, Bob cannot determine whether Charlie is genuinely expressing his emotional state or concealing it. As a

---

object or letter reinforced her learning. This is similar to how Alex benefits from direct feedback in her sensory acquisition process. Prior to Samuel Gridley Howe's intervention, Bridgman was isolated from the outside world due to her disabilities. Learning the manual alphabet and the use of raised labels enabled her to communicate and overcome her isolated condition. Another example is Helen Keller, who lost her sight and hearing due to an illness at a young age. Her teacher, Anne Sullivan, utilized techniques similar to Howe's, including the manual alphabet, to communicate with Keller and instruct her in reading, writing, and speaking. Sullivan frequently poured water over Keller's hands while tracing the letters of the manual alphabet with her other hand, aiding Keller in forming connections between objects, words, and ideas. After beginning her education with Anne Sullivan, Helen Keller made rapid progress in learning. She went from being unable to communicate meaningfully to learning how to use the hand alphabet, read Braille, and even speak. Her ability to learn several languages, including French and German, and to write books, give public speeches, and graduate from college, is evidence of her remarkable cognitive development. Another example is the Tadoma method, a communication method used by some deaf-blind individuals. The method involves the deaf-blind person placing their hands on the speaker's face or throat to feel the vibrations and movements associated with the spoken word. Tadoma was originally developed for Olaudah Equiano (also known as Gustavus Vassa), who became deaf and lost the ability to speak following an illness. This method enabled him to continue to communicate effectively with others. In all three cases, the introduction and use of alternative modes of communication resulted in significant improvements in access to information, social interaction, and the ability to express oneself. This indicates a strengthening of cognitive abilities, which has been documented and analyzed in educational and psychological studies. These improvements are not only anecdotal but have been confirmed to be of vital importance in the cognitive development of individuals with sensory disabilities through multisensory integration.

result, Bob's ability to fully comprehend and interpret Charlie's mental state is significantly restricted. This leads to a superficial and potentially inaccurate form of interaction, in which mutual understanding is limited. Communication is often reduced to a mere exchange of verbal statements, resulting in a loss of emotional and psychological depth. While Bob and Charlie may develop language that is rich and complex in content, their ability to infer hidden feelings, intentions, or thoughts remains extremely limited.

Let's emphasize here the concept of managing conflicting information (the non-redundant signals of Section 2.3) in multimodal communication, which involves the existence of some sort of multimodal syntax, pragmatics, and syntactics. Although briefly mentioned, the concept of multimodal syntax, pragmatics, and syntactics will be useful in Chapter 5. Consider the example of a communicative exchange between Bob and Charlie where verbal and nonverbal information conflict. Bob claims he is not cold, yet his voice and body shake. While Bob's verbal message suggests he is not cold, his nonverbal signals indicate otherwise. The difference between the words spoken and the body language displayed creates an interpretive dilemma for Charlie. He must determine which information is more reliable or truthful. To solve this dilemma, Charlie must rely on his prior experience. This experience includes not only his knowledge of Bob's past behaviour but also general situations in which he has observed shaking as an indicator of cold or other emotions or conditions. Charlie must evaluate and prioritize information, considering both Bob's verbal statement and his body language. This process requires the use of multimodal pragmatics, which involves evaluating the context, past experiences, and likely causes of the observed behaviour to determine which information is more significant. Additionally, Charlie should apply multimodal syntax to organize and interpret information based on its coherence. For instance, Bob's voice shaking could be combined with his physical shaking to suggest that Bob is cold, despite his words. Multimodal syntactics refers to how Bob and Charlie construct their messages, considering both verbal and nonverbal channels. Effective communication requires verbal and nonverbal elements to support each other in conveying a coherent message. However, incongruent situations, such as the one exemplified, pose a challenge for Charlie in decoding and correctly interpreting conflicting messages. Information from different sensory channels should not simply be added together. The understanding of such information requires complex interpretation involving experience, context, and intelligent management of the discrepancy between different channels.

When Charlie is faced with a discrepancy between Bob's words and his body language, an evaluation process is triggered that can operate on two distinct levels: (i) at the unconscious level. Many of our interpretations of language and nonverbal behaviour occur automatically and without our explicit recognition. For instance, Charlie's brain may involuntarily perceive Bob's trembling as a cold signal, despite Charlie consciously processing Bob's verbal words stating otherwise. This ability to quickly and automatically process nonverbal signals is crucial in everyday communication and plays a significant role in our ability to respond appropriately and intuitively to social situations. (ii) Additionally, at the conscious level. At the same time, Charlie could engage in a more intentional and conscious analysis of conflicting information. He could actively consider the possible truthfulness of Bob's words, taking into account the current context and his past experiences with Bob. This level of conscious processing enables him to form a more reflective and thoughtful judgment about the situation. The integration of conscious and unconscious processing levels provides a more complete understanding of the communication process. The unconscious level facilitates quick and intuitive responses, while the conscious level allows for a more thorough and reasoned analysis of complex situations.

As theoretically supported by the thought experiments of Alex, Bob, and Charlie, the phenomenon of cross-modal binding is deeply connected to language, communicative systems, and ToM. The first experiment explores the development of consciousness, language, communication, and thought in a being named Alex who lacked initial sensory perceptions. It investigates how the emergence of cross-modal binding, and the progressive acquisition of senses, directly influence these abilities. Alex, initially in a state of nonconsciousness, begins to develop a sense of self and rudimentary categorization skills with the acquisition of hearing, vision, and eventually, through multisensory integration, progresses toward the formation of both language and communication and more complex consciousness. In the second part, we examine the interaction between Bob and Charlie in a context of communication limited to a single-component verbal channel, devoid of nonverbal nuances, to determine how this restriction impacts ToM. Bob's inability to access nonverbal clues significantly limits his understanding of Charlie's mental states, emphasizing the importance of multimodal communication for complete mutual interpretation and understanding. In the second example of Bob and Charlie, we see how Charlie is required to arrange the meaning of non-redundant signals hierarchically, following a set of rules to infer the correct truth value of the message within a multimodal communication. The concept of hierarchically organized symbol systems and truth values will be further explored in Chapter 5.

3.4 Neuroscientific implications of Multimodal mind theory

For the multimodal mind theory to be considered valid, studies in the field of neuroscience should demonstrate a series of consequential phenomena. Six main areas of investigation can be identified in relation to humans: (i) Brain regions involved in cross-modal binding. Brain regions responsible for cross-modal binding, such as the superior temporal sulcus, posterior superior temporal sulcus, and temporo-parietal junction, are expected to be actively involved in both ToM and the processing of language or multimodal signals. These areas should show high activity during tasks that involve understanding others' mental states or comprehension of language, especially when such tasks require greater integration of multiple sensory inputs. (ii) Functional connectivity between brain areas. Significant functional connectivity is expected between areas involved in ToM (such as medial prefrontal cortex, temporo-parietal junction, and anterior cingulate cortex) and areas involved in language (including Broca's area, Wernicke's area, and related subcortical structures). Tasks that require the use of language to infer others' thoughts or feelings could show such connectivity, suggesting the existence of a network that supports both ToM and language through cross-modal connections. (iii) Developmental studies. Developmental studies should reveal that progress in cross-modal binding skills is associated with improvements in both ToM and language. Neuroimaging studies can confirm this by monitoring the maturation of brain networks involved in cross-modal binding, language, and ToM over time. (iv) Effects of training and neuroplasticity. Interventions aimed at stimulating cross-modal binding should potentially affect language and ToM. Pre - and post - intervention neuroimaging could reveal changes in areas specifically targeted by training, as well as in regions associated with the ToM and language domain, suggesting a common neural basis. (v) Intercultural and bilingualism studies. Interventions aimed at stimulating cross-modal binding may potentially affect individuals who are bilingual or multilingual, or those from cultures with different modes of sensory involvement. These individuals may exhibit different patterns of neural activation or connectivity during tasks related to ToM, suggesting that the development of ToM and language skills is modulated by the type and richness of sensory experiences. (vi) Populations with atypical development. Abnormalities in brain regions or networks implicated in cross-modal binding could be observed in groups with atypical development, such as individuals with ASD who may have difficulty with ToM or those with language disorders. Comparing these groups to neurotypical controls could reveal the neural basis of ToM interaction, communicative systems, and cross-modal binding.

The implications of the relationship between ToM, communication, and cross-modal binding apply not only to humans, of course, but also to the rest of the animal kingdom. Nonhuman species may exhibit mechanisms similar to those of ToM, language, and cross-modal binding in humans, adapted to their cognitive and communicative abilities. Three main considerations arise in this context. (i) Brain Areas Related to Social Cognition: Nonhuman species likely have brain areas related to social cognition that function similarly to those involved in sensory integration and Theory of Mind (ToM) in humans. These areas may include regions involved in multimodal communication and social interaction, which exhibit similar functionality to areas of human language. (ii) Characteristics of Animal Communication Systems: Analysis of animal communication systems may reveal elements comparable to human language, such as compositionality, recursiveness, and the ability to exploit hierarchically complex symbol systems. This notion is explored in more detail in Chapter 5. Species with unique social or sensory systems are also considered. Examining species with atypical social systems or unique sensory modalities, such as eusocial insects or bats that use echolocation, could provide insights into how different sensory and social systems support complex ToM-related cognitive processes. These observations could expand the understanding of the neural and functional basis of social cognition beyond the human model.

After reviewing the literature on each of the aforementioned areas, it is evident that current studies only partially support the hypothesis of a relationship between ToM, language, and communicative systems, as well as cross-modal binding, in both humans and non-human animals. However, there are still numerous unanswered questions and some inconsistent results. (i) Research on cross-modal binding in humans is well-developed. Numerous studies have shown how the brain integrates information from different sensory modalities to form a coherent perception of reality. These studies support the idea that cross-modal binding is critical for higher cognitive tasks, including language and ToM (Calvert, 2001; Saito et al., 2005, Hocking & Price, 2008; Lin et al. 2018; Deng et al., 2013; Hollenstein et al., 2021; Rogers, 2023;). (ii) Research on the functional connectivity between brain areas involved in ToM and language shows promising results, with neuroimaging studies showing overlapping neural networks for these functions. However, the interactions between these networks are complex and require further investigation for full understanding (Veroude et al., 2010; Gaudet et al., 2020; Wade et al., 2018; Balgova et al., 2023; Paunov et al., 2022). (iii) Longitudinal studies on the development of ToM and language skills in relation to cross-modal binding are similar. Studies by Ebert (2020), Białecka et al. (2023), Grazzani et al. (2018), Wang et al. (2014), and Zaho et al. (2021) support this idea. (iv) Studies on the effects of training and therapeutic intervention on language skills and ToM show that it is possible to induce neuroplastic changes that improve these abilities, suggesting some overlap in the neural networks involved (Ku & Sung, 2022; Tucci et al., 2016; Kasuya-Ueba et al., 2020). However, the extent and specificity of these effects vary widely among studies. Research on bilingualism and cross-cultural differences in language processing and ToM is extensive and shows that these experiences can affect brain structure and function. However, the results are sometimes mixed, and the relationship between bilingualism, ToM, and cross-modal binding needs further clarification (Xia and Haas, 2023; Marinis, 2018). (vi) Studies of individuals with atypical development, such as those on the autism spectrum, have provided important insights into the neural basis of ToM and language (Larson et al., 2023; Rossello et al., 2020; Bulgarelli et al., 2022). These studies suggest differences in the connectivity and activation of certain brain areas, but the picture is complex and not fully defined. Regarding studies conducted on non-human animals, there are many behavioural and neuroscientific studies that explore how animals integrate sensory information in social contexts. (i) These studies support the idea that cross-modal binding-like abilities are present and important for social cognition in different species. (ii) Neuroscientific research on the neural basis of social cognition has been growing, with studies identifying brain areas involved in complex social behaviours (Lee et al 1993; Adolphs, 2009; Traniello, 2021). However, it is difficult to directly compare these areas with those in humans due to anatomical and functional differences between species. (iii) Studies on animal communication systems have shown considerable variety and complexity. However, equating these systems with human language remains controversial. Animal communication is recognized as crucial to the negotiation of social relationships and may perform similar functions to language in expressing internal states.

In the next chapter, the epistemological problem will be addressed by showing how the construction of protocols using the multimodal shift mechanism can provide more accurate results than currently existing experimental paradigms.

4 The logic problem and its possible solutions

The previous chapter discussed the concept of multimodal shift as a potential indicator of ToM in non-linguistic animals, and also pointed out that not all forms of multimodal shift can be used to assess ToM. To serve as a valid indicator of ToM, multimodal shift must have specific features, including that it does not occur as an invariant response to specific stimuli, that it uses redundant signals, and that it occurs in the presence of potential perceivers. Multimodal shift may be a potential indicator of ToM in non-linguistic animals and can therefore be used to design experiments to test ToM in non-linguistic animals. To support the argument that non-linguistic animals can exhibit ToM, multimodal mind theory has been proposed to disentangle the relationship between ToM and language. The development of both abilities is thought to depend on the development of cross-modal binding, i.e. the ability to integrate information from different sensory modalities. Several research studies support the idea that cross-modal binding underpins the development of language and communication systems as well as ToM in both humans and other animals. This theory, termed multimodal mind theory, could reconcile the different schools of thought on the relationship between language and ToM, suggesting that the ability to process and integrate sensory information is critical for the development of both. Two thought experiments were presented (section 3.3) to support the hypothesis that cross-modal binding is crucial to the development of both language and ToM. The first thought experiment considers a hypothetical creature called Alex, who initially lacks any form of sensory perception and, consequently, consciousness or thought. As Alex acquires sensory abilities, such as hearing and sight, a rudimentary sense of self and an understanding of the surroundings begins to develop. The cognitive development of Alex begins with the gradual acquisition of sensory capabilities and the ability to differentiate self-produced sounds from external sounds. More complex thought processes, categorization, and the emergence of language and consciousness are based on the integration of multisensory information at the unconscious level. This thought experiment suggests that unconscious integration of sensory input may be essential for the construction of reality and the development of higher cognitive functions. The second thought experiment involves two individuals, Bob and Charlie, in a hypothetical world where communication is limited to a single-component, unimodal verbal channel. This unimodal and single component communication condition severely limits Bob's ability to understand Charlie's true emotional state, as all nonverbal cues that typically accompany communication, such as tone, facial expressions, and body language, are absent. This scenario supports the idea that ToM may rely on multimodal communication. Without access to multimodal signals, understanding the intentions, feelings or thoughts of others is significantly impaired, leading to superficial and potentially misleading interaction. The chapter concludes with a discussion of the potential implications of multimodal mind theory from a neuroscientific perspective. Finally, in this chapter we will see how the multimodal shift phenomenon can be used to overcome the epistemological problem of the dual interpretability of data. Experiments aimed at testing the presence of ToM in non-linguistic animals face this problem. Section 4.1 provides a detailed explanation of this problem, including its origin and development. Section 4.2 outlines a solution to the logic problem when testing a non-linguistic animal's ability to understand another subject's perception. Section 4.3 explains how to overcome the logic problem when testing a non-linguistic animal's understanding of another subject's false beliefs. Section 4.4 presents preliminary data from a pilot experiment conducted on two elephants at the Rome Zoo. Although the data are still preliminary, they are worthy of reporting within this research.

4.1 The history of the issue of discriminating between mindreading and behaviour reading behaviours.

Consider the study done by Povinelli and Eddy (1996) on a chimpanzee. The chimpanzee must choose between two people to obtain food. One person can see the animal, while the other is blindfolded and unable to see. The animal must interact with one of the two people to obtain food. If the animal consistently chooses to interact with the person who can see it, this behaviour can be interpreted in two ways. On the one hand, the animal can be interpreted as comprehending the difference in the mental states of the two individuals. In other words, the animal engages in reasoning of the following type: (i) People see because of their eyes; (ii) One of the two people has their eyes covered, and (iii) therefore cannot see me; (iv) the other, however, has open and unobstructed eyes and can see me; (v) In order to obtain food, I must interact with the person who can see me. The interpretation of animal behaviour as understanding the mental states of the two individuals means that the animal attributes the ability to see to other subjects and relies on this to choose how to behave. Alternatively, another interpretation can be argued: the animal does not attribute any mental states to people, but instead relies on simple visual signals (the presence or absence of a blindfold) to predict which people are more likely to give it food. The animal has learned, during several repetitions of the test, to associate the presence of the blindfold with a lower likelihood of obtaining food, while the opposite is true when the blindfold is absent. This demonstrates a mental understanding of cause-and-effect, but not of the other's mental state. The effect of the blindfold is the inability to obtain food from this person. In the first case, we will refer to this interpretation as the mind-reading hypothesis. In the second case, we will refer to it as the behaviour-reading hypothesis. The mind-reading hypothesis holds that an individual's prediction of other behaviours is based on attributing mental states to others (ToM). These attributions are inferred from observed behaviours, environmental contexts, and the understanding that certain mental states typically lead to the performance of certain behaviours. The behaviour-reading hypothesis suggests that predictions are made based on observed behaviours and environmental contexts, rather than mental states. The ability to predict others' behaviours is achieved through learning the cause-effect association between a specific situation and the resulting behaviour.

The question then becomes how it is possible to distinguish the mind-reading hypothesis from the behaviour-reading hypothesis empirically, by means of experimental tests. This challenge is known as the logical problem or the problem of observables (Povinelli & Vonk, 2006; Lurz, 2011; Heyes 2014). The two hypotheses are complementary and cannot be experimentally distinguished from each other by any of the existing experimental protocols. The two hypotheses are based on the same data and observations. Whether the animal in the previous example predicts human behaviour through behaviour-reading or mind-reading, it is still relying on the same observed behaviour (one of the two people is blindfolded) and the same environmental contexts (it has to interact with one of the two people in order to get food). Designing experiments that can incontrovertibly determine which of the two mechanisms is relied on by a given individual is difficult due to the overlap of data. However, this difficulty does not stem from experimenters' inability to design valid experiments, but rather from the nature of the experiments themselves. Any positive result in an experiment designed to test the presence of ToM that is based on correlations between observable traits and behaviours exhibited by others can always be interpreted in terms of a complementary behaviour-reading hypothesis[59], and not necessarily as a mind reading ones[60].

To be clear, proponents of the logic problem claim that there is a complementary behaviour-reading hypothesis for every mindreading hypothesis. The mindreading hypothesis suggests that participants excel in mindreading tasks because they can anticipate an agent's actions by understanding the agent's mental state. For example, the chimpanzee in Povinelli and Eddy's (1996) experiment might reason that (i) for a person to give me food, he must first be able to see me ask him for food (ii) one of the two people can see me (mental states) (iii) then that person will give me food (behaviuor). On the other hand, a complementary behaviour-reading hypothesis suggests that participants excel in a Theory of Mind tasks not by understanding the agent's mental state, but by predicting their behaviour solely based on observable patterns. In this case, the chimpanzee in Povinelli and Eddy's (1996) experiment might reason that (i) a person tends to give me food when the trajectory from his eyes to me is clear as I ask him for food (ii) the trajectory between me and that person's eyes is clear (situation), so he is likely to give me food (behaviour). Proponents of the logical system argue that any current experimental protocol used to test ToM in non-linguistic animals cannot discriminate between a mindreading and a complementary behaviour-reading hypothesis. To put it differently, for someone to understand and predict another person's actions based on their mental states (ms) in a given situation (s), they must recognize the regularity that links the situation, the cognitive state, and the behaviour b (s -> ms -> b). But if individuals already know the pattern that connects the situation directly to the behaviour without considering the cognitive state, why complicate matters by adding the step of inferring the cognitive state?

---

[59] A complementary behaviour-reading hypothesis and overlapping data are not always possible and present. In some cases, experiments can be interpreted using either the mind-reading or behaviour-reading hypothesis, but these interpretations are not always complementary. Discriminant testing can be conducted to determine which hypothesis is valid. If the behaviour-reading hypothesis is not complementary, it is referred to as a minimal behaviour-reading hypothesis. In this case, a subject uses observable data to predict another individual's behaviour, which may differ from the data used in ToM activities. Therefore, if we are dealing with a minimal behaviour-reading hypothesis, the logic problem does not arise.

[60] The issue of dual interpretability of data is closely related to the question of simplicity. In the field of scientific research on animal cognition, the Morgan canon has historically been applied. As explained in chapter one, Morgan's canon is simply the application of Ockham's razor to the field of cognitive science. According to the canon, when all things are equal, we should always choose the simplest psychological explanation for animal behaviour. The question is whether the theory of mind explanation is simpler than the complementary reading of behaviour. Some researchers, such as Povinelli and Vonk, support the hypothesis that behaviour reading is simpler, while others, such as Call and Tomasello, believe the opposite.

Advocates of the logical problem question the need for this additional step, especially considering that current experiments on mindreading don't provide a clear reason to do so.

To solve the logic problem, it is necessary to design experiments that do not have the same characteristics as those used so far to investigate ToM. In order to illustrate how to design such experiments and to demonstrate that an experiment that exploits the phenomenon of multimodal shift is the most suitable for testing ToM in non-linguistic animals, existing protocols will first be presented, and then their common features will be discussed. Finally, a potential solution to the logic problem proposed by Lurz will be presented and evaluated and it will be shown how the solution using multimodal shift is more effective in testing ToM than Lurz's solution.

Early theoretical foundations regarding the issue of discriminating between complementary hypotheses[61] was initially identified by philosophers Harman, Dennett, and Bennett in 1978. They noted that the divergent positions on the interpretation of experimental results on ToM were a natural effect of the way experiments were and are designed. In response to Premack and Woodruff's 1978 study, Dennett proposes a system for categorizing the intentionality or mental states that a subject may have. The categorization system comprises three levels. (i) The level zero intentionality consists of having no mental states, no beliefs or desires, and consequently no intentionality, intelligence, or form of communication. For instance, a thermostat reacts to temperature changes but does not possess the capacity to 'believe' or 'desire' a specific state. The thermostat operates solely based on preset mechanisms and lacks the capacity to interpret or comprehend information in a meaningful way. Unlike living beings with more complex nervous systems, it cannot process information or make decisions based on contextual understanding. (ii) Level one intentionality consists of having mental states such as beliefs and desires, but not other mental states. According to Dennett, non-human animals such as birds or rodents may have beliefs and desires, but they may not have the ability to understand or represent that other animals may have the same beliefs or desires. For instance, a pigeon may believe that there is food in a particular place and have a desire to eat it but it does not necessarily have the capacity to understand or represent to

---

[61] Harman, Dennet and Bennet never use the term "complementary hypotheses" nor the term logical problem. However, I believe that the earliest traces of the logical problem can be traced back to them.

itself that another pigeon might have the same belief or desire about where the food is. These animals may have complex interactions with their environment and respond to specific stimuli. However, they do not show evidence of the ability to reflect on the beliefs or desires of other members of their species. (iii) The level two intentionality involves the ability to represent other mental states as mental states. This level corresponds to ToM. The issue of discriminating between complementary hypotheses can also be interpreted as the challenge of determining whether an animal possesses a first or second level of intentionality, according to Dennett's categorization system.

The issue of discriminating between complementary hypotheses is emphasised by Povinelli and Vonk in their well-known 2006 paper. Povinelli argues that in order to solve the logic problem (later named as such in the literature by Hurley & Nudds 2006, Lurz 2009, and Halina 2015), an experimental paradigm must be designed that a subject can only pass if they can accurately predict another individual's behaviour through ToM. If the subject relied on a behaviour-reading mechanism, the test would fail. Cecilia Heyes (1998) previously addressed the challenge of constructing such an experimental paradigm. She argued that all existing experimental designs at the time were ineffective in solving the logic problem and proposed a new experimental paradigm, the experience projection (EP) paradigm, inspired by another paradigm, the guesser-knower (GK) paradigm developed by Povinelli (1990). Let us now briefly examine the functioning of the GK protocol, after which we shall proceed to analyze the EP protocol. The GK protocol requires subjects, which may belong to different species, to face a pair of experimenters: one who knows the location where food is hidden (the knower) and one who does not (the guesser). The experimenters indicate, respectively, which container contains the hidden food, correctly and incorrectly. The subjects' task is to choose the container that contains the food. There are two ways in which subjects can solve this task: either by inferring the experimenters' mental states - i.e. by determining who actually knows where the food is and who is guessing (mind-reading hypothesis) - or by applying learned rules based on surface clues to determine which experimenter is most likely to indicate the correct location (complementary behaviour reading hypothesis)[62].

---

[62] The 'guesser-knower' paradigm has been utilized in various forms to investigate social cognition in several species, such as chimpanzees, rhesus monkeys, dogs, and crows. However, due to its reliance on a cooperative-communicative context, which is particularly relevant to humans but not necessarily to all other

A practical example of the GK protocol is that of Catala et al. (2017). Catala conducted a GK test with dogs to explore their ability to discriminate between human informants based on their visual access to events, specifically in the context of detecting hidden food. The study was conducted in a room equipped with a three-camera video recording system. Opaque containers and a removable screen were used to manipulate the visibility of the food concealment process (see Figure 1). Prior to the main tests, the dogs underwent a pre-training phase to familiarise themselves with the test environment and task. This phase was conducted in a manner that avoided any potential bias on the part of the dogs, with regard to the containers or informants that might be involved. The pre-training phase involved a series of steps to gradually increase the number of containers and introduce a screen to block the dog's view during the food concealment process. At first, only one container was presented to the dog, and the food was visibly concealed. Subsequently, the number of containers was increased to four, and measures were taken to mask odor and sound during food concealment, making all containers appear identical. Following pre-training, three main tests were conducted using the Guesser-Knower paradigm. The person who hides the food in the containers is called the Knower and the other person the Guesser. The first test was called the Guesser Present (GP) test. (i) Guesser Present (GP) test: a bait was placed in two of the four containers, rather than just one, while the dog was unable to observe the food being concealed. In fact, the dog's view was momentarily obstructed by a screen. After the screen was removed, both the Knower and the Guesser pointed to one of the container with food, both having observed food concealment. The purpose of this test was to determine whether the dog had a preference for a specific informant or other types of bias. (ii) Guesser Absent Test (GA): This test aimed to assess the dog's ability to discriminate between two informants based on their presence during food concealment. Before the food was hidden in the containers, the 'Guesser' informant would leave the room and return only after the baiting process had finished. Afterwards, the 'Guesser' would indicate an empty container, while the

---

species, this paradigm has sparked considerable debate regarding its ecological validity in experimental design. For instance, chimpanzees may not perform well in experiments that require them to interpret communicative gestures, such as pointing, since they rarely engage in such communication in the wild. In contrast, human children and domestic dogs have developed a specialized sensitivity to such cooperative communicative cues. Therefore, while the 'guesser-knower' paradigm remains a popular method for studying social cognition in various species, it is important to interpret cross-species differences in performance by taking into account the ecological validity of the task for the specific species being studied. This requires assessing whether the task is natural or relevant to that species.

'Knower' would indicate the correct one. (iii) Guesser Looking Away (GLA) Test: In this test, a third person (the 'baiter') was introduced to assess the influence of food handling in addition to the two informants. The 'baiter' hid the food in one of the containers while the informants looked in fixed, predefined directions, simulating different visual accesses to the action of the 'baiter.' Only the 'Knower' had the ability to see the concealment, while the 'Guesser' did not. In this test, the baiter is positioned between the guesser and the knower, hiding the food, but not pointing to any of the boxes. The aim of all tests was to observe the dogs' choices and determine their ability to discriminate between informants based on their visual access to food concealment events. The tests were designed to minimize other cues such as smells or sounds. The study results indicate a significant dog's preference for the container indicated by the Knower in two out of three tests. Specifically, in the GA test, the dogs select the container indicated by the Knower 72.3% of the time, well above the expected 50% chance level. Similarly, in the GLA test, the dogs select the container indicated by the Knower 61.7% of the time. In the GP test, although the dogs select the container indicated by the Knower was 56.2%, it was not statistically significant, but still showed a trend towards a preference for the container indicated by the Knower. This shows that the dogs were able to distinguish between a person who knew where the food was located (the Knower) and one who did not (the Guesser) and preferred to follow the suggestions of the one who knew the location of the food.

As mentioned just above, in the GK protocol, the dog might perform reasoning such as: (i) a person to point me to where the food is must have seen where the food was hidden (ii) the knower has seen where the food is hidden therefore knows where the food is (ms) (iii) therefore the knower will correctly point me to where the food is (b). At the same time, the dog might carry out a different reasoning: (i) a person indicates to me with more probability of certainty where the food is if the trajectory from his eyes to the food is clear (ii) the knower had the trajectory from his eyes to the food clear (s) (iii) therefore if I follow the knower's directions I will be more likely to find food (b). In this case, the dog could correctly predict behaviour based on a complementary behaviour-reading hypothesis (s -> b). Again, as already explained, if two complementary solutions are available, there is no way to discriminate which of the two is true. The GK protocol cannot consequently solve the logic problem. This ineffectiveness in solving the logical problem is noticed by Heyes, who tries, based on the GW protocol, to design a protocol that overcomes the GW methodological problems.

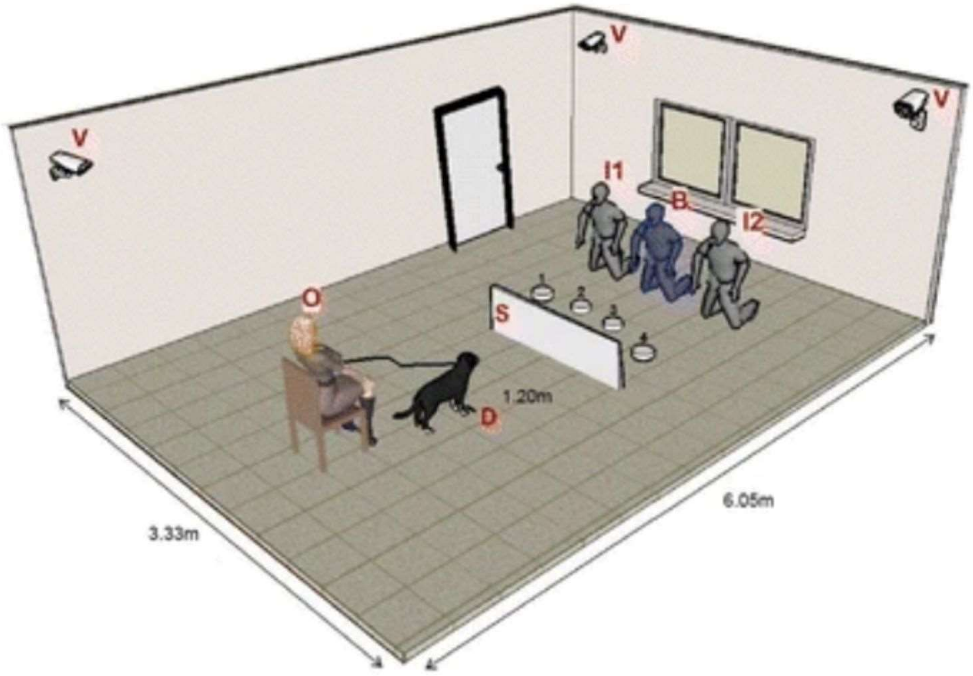


Figure 1 shows a sketch of the testing room, including the positions of three video cameras (V), the owner (O), the dog's releasing point (D), the screen (S), the four containers (1, 2, 3, 4), the two informants (I1 and I2), and the baiter (B) in blue, who was only present in the GLA condition. This figure is reprinted from Catala et al. 2017.

Hayes' EP protocol differs from the GK protocol in that it is based on subjects' responses according to their past experiences. The protocol consists of two experimental conditions: (i) a condition in which subjects have experienced being in the position of the observed agent. In this condition, subjects have directly experienced a situation similar to the one they are observing. Due to their direct experience, individuals can project their own mental states onto the observed agent and interpret their behaviours based on what they themselves would have seen or perceived in that situation. In contrast, when subjects have not had a similar experience, observed without having experienced it. Their ability to accurately interpret the agent's mental state may be limited or absent because they cannot draw on relevant personal experiences. Heyes illustrates this experimental protocol using a hypothetical experiment known as the goggle experiment (Figure 2). The experiment is very similar to the one presented in the introduction of this chapter. The hypothetical goggles experiment illustrates this idea. During the experiment, a chimpanzee was observed as it learned to request food from the experimenter who could see it, while ignoring the one who could not. This allowed the participants to differentiate between the two experimenters based on their observed behaviours. Next, the chimpanzees tried on two pairs of goggles: a blue pair with clear lenses that allowed them to see, and a red pair with opaque lenses that prevented them from seeing. This direct experience was crucial because it allowed the chimpanzees to understand the effect of the glasses on vision based on their own sensory experience. After this experiment, the experimenters will wear either clear blue or opaque red goggles, and the chimpanzees will be given the choice of which experimenter to "ask" for food. It is expected that participants who have experienced the goggles and can attribute the ability to 'see' or 'not see' to the agent will ask the experimenter wearing the transparent goggles for food, rather than the one wearing the opaque ones. A control group without direct experience with the glasses would not need to make this distinction, as they lack a reference point based on experience to interpret the agent's mental state. If experienced participants showed a significant preference for the experimenter wearing the clear goggles, this would suggest that they are using their own past experience to attribute mental states to the experimenter, thus demonstrating a form of ToM according to Heyes.

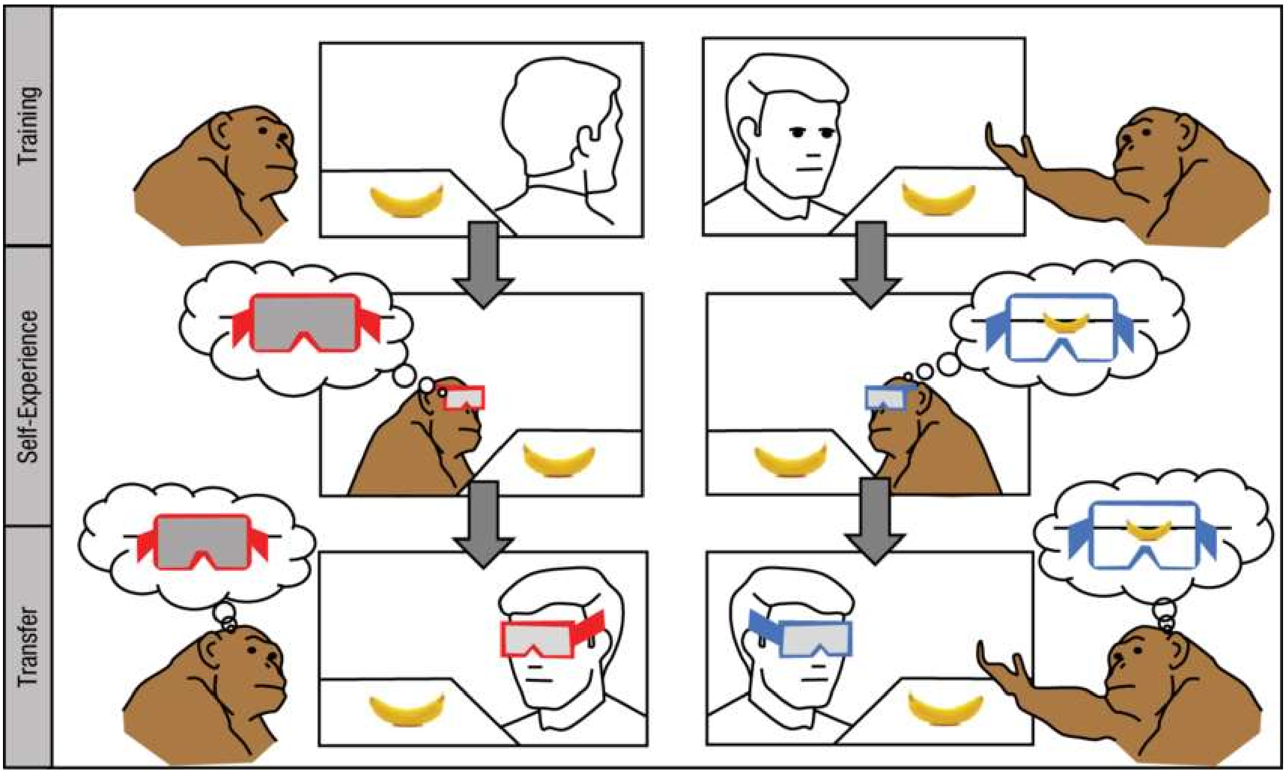


Figure 2 shows Heyes' (1998) original description of the experience projection method, reprinted from Lurz and Krachun (2019).

Heyes' work on the EP protocol was purely theoretical. Several applications inspired from his work have come to exist, including Dally et al.'s (2004) work on scrub jays and Stulp, Emery, Verhulst, and Clayton's (2009) work on scrub jays, which will be discussed in more detail later in this chapter. Heyes believes that his EP protocol can solve the logic problem, but Robert Lurz (2009, 2011, 2019) disagrees. Lurz notes that existing protocols for testing the attribution of cognitive states in animals, including Heyes', do not solve the logic problem. The question remains whether the subjects, such as the chimpanzees in Heyes' hypothetical goggles experiment, are acting with ToM, or whether a complementary behaviour-reading hypothesis is possible instead. In Heyes' experiment, Lurz points out that chimpanzees may learn from wearing glasses that they can obstruct direct contact between their eyes and the observed object. Perner (2012) later confirms this observation. Based on prior experience, when an observer's line of sight to an object is obstructed, they may not interact with the object. In this case, the observer may rely on the current stimuli, such as the goggles being opaque and blocking the experimenter's line of sight to the food, to prefer interacting with the experimenter wearing transparent lenses. This preference is based on the observer's prior experience rather than attributing the ability to see to the subject.

To summarize, both the GK and EP protocols are ineffective in resolving the logic problem. However, this is not the only protocol that has been found to be ineffective. All other protocols that have been employed to assess ToM without the use of language have also been shown to be ineffective in resolving the logic problem. When analyzing the protocols used to test ToM or some level of its development in non-linguistic animals, it is possible to divide the various experiments into three main groups (Lurz, 2011; Heyes 2014). (i) Those designed to test the ability to attribute perceptual abilities to another individual. (ii) Those designed to test the ability to attribute knowledge states to another individual. (iii) Those designed to test the ability to attribute beliefs to another individual. To illustrate the characteristics of these three categories of experiments, we will analyze the most well-known experiments. We will demonstrate how each of them, following the observations proposed by Lurz (2009, 2011, 2019) and Perner (2012), is unable to solve the logic problem in question. The identification of the common features of these experiments will reveal the reason for their inability to solve the logic problem. Consequently, this understanding will inform the design of subsequent experiments, ensuring that they do not exhibit these characteristics and thus enabling them to successfully address the logic problem.

4.1.1 Perception attribution experiments

In the first category of experiments, it is important to mention the experiment conducted by Hare and colleagues (2000) on chimpanzees. The researchers conducted a series of experiments in which they tested two chimpanzees, one dominant and one subordinate, against each other for access to two pieces of food. The objective of the study was to evaluate whether subordinate chimpanzees could comprehend and utilize their dominant conspecifics' visual limitations to gain an advantage in competitive food situations (Figure 3). The experiment consisted of multiple phases, with varying food visibility conditions. In certain scenarios, both chimpanzees had unobstructed visual and physical access to the food. As anticipated, in these scenarios, the dominant chimpanzee tended to consistently obtain the food due to its superior hierarchical position. In other arrangements, the food was placed in a location visible only to the subordinate chimpanzee, not the dominant. This was a critical test to determine if the subordinate could utilize its exclusive knowledge to acquire the food without engaging in direct conflict with the dominant. To eliminate the possibility that the subordinate chimpanzee was simply monitoring the dominant's behaviour and avoiding the food that the dominant was heading for, the experimenters introduced a variation. The subordinate chimpanzee was given a small time advantage, which forced it to choose which food to reach for (the one visible to both or the one visible only to it) before the dominant was released into the area. This approach ensured that the subordinate's choice was based on the dominant's visual situation, rather than an immediate reaction to the dominant's movements. The results showed that subordinate chimpanzees were often able to successfully obtain food in situations where they had a visual advantage. This suggests a sophisticated understanding of what their dominant conspecifics could or could not see, according to Hare and colleagues.

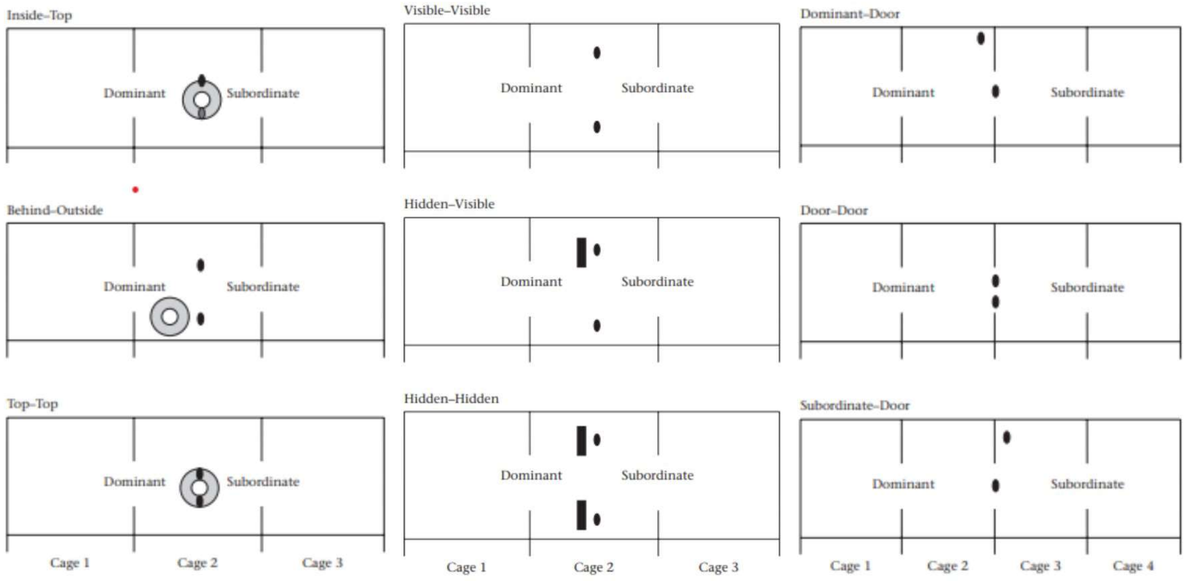
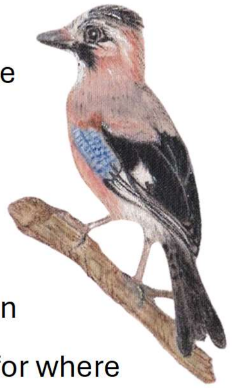


Figure 3 displays the test conditions to which the chimpanzees were exposed, as shown in the three graphs. This information was reprinted from Hare et al. 2000.

Hare and colleagues' interpretation can be compared to their complementary behaviour-reading hypothesis. This alternative perspective suggests that chimpanzees may have learned through experience which strategies lead to success in certain social and environmental configurations, rather than demonstrating an understanding of others' perceptions. From this perspective, subordinate chimpanzees may be guided by simple behavioural rules based on their observation of the consequences of their own past actions, rather than by a sophisticated understanding of the mental state of the dominant conspecific. For instance, chimpanzees may have learned that approaching food that is not visible to the dominant reduces the likelihood of a confrontation. However, they may not understand that this correlation is due to the dominant's lack of visual perception. Additionally, chimpanzees may use social and behavioural cues, such as the dominant's gaze direction or body posture, to assess the situation without explicitly attributing 'vision' or 'non-vision' to the conspecific. This type of situational and behavioural learning differs from ToM in that it does not require the animal to understand the perceptions of others, but only to recognize and respond to observable patterns of behaviour. The experimental paradigm, and others that follow the same mechanism, fail to overcome the logic problem.

One noteworthy experiment is the study conducted by Dally et al. (2004) on scrub jays. The study aimed to investigate the caching strategies employed by these birds to protect their hidden food from being stolen by other members of their species. The experiment was conducted with two main observation conditions during the caching[63] phase: a private condition, where no conspecific observers were present, and a public condition, where other conspecifics could observe their behaviour. The scrub jays were then offered food, which could be hidden in two areas - one well-lit and the other in shade (see Figure 4). When there were no observers present, the scrub jays did not show a preference for where to hide their food, selecting both well-lit and shaded areas equally. However, when observed by other scrub jays, they tended to choose shaded areas to hide their food. This behaviour is likely an attempt to reduce the chance of the observers remembering the location of the hidden food. Approximately three hours after caching, the food-containing trays were returned to the scrub jays under private conditions, allowing them to retrieve the supplies without observers. The scrub jays that had hidden food in the private condition showed no preference for retrieving food from either the illuminated or shaded areas. In contrast, the scrub jays that had hidden food in the public condition tended to retrieve food located in well-lit areas first, rather than shaded areas. This suggests that they considered the shaded areas to be safer and therefore prioritized maintaining the hidden stores. The researchers' interpretation is based on the 'mind-reading' hypothesis, which suggests that scrub jays attribute the ability to see and remember to their conspecifics. Under public conditions, shaded areas are perceived as less visually accessible to observers, making them preferable for safely hiding food. However, in the absence of observers (private condition), this preference disappears as there is no risk of others seeing where the food is hidden. In a subsequent experiment, the brightness variables were substituted with the distance from the observer, yielding comparable results: scrub jays chose to conceal food in locations that were more challenging to detect when other members of the same species were present.



---

[63] Caching is a behaviour observed in many animal species where they hide food or objects for future use. This behaviour is particularly common in species that face periods of food scarcity or difficulty in finding food. Animals such as jays, crows, squirrels, and some rodents hide food in various locations that they can remember and return to later. Caching is a complex behaviour that requires the ability to plan ahead and remember the location of hiding places. Remembering the location of each hidden piece of food requires significant cognitive skills, especially spatial memory.

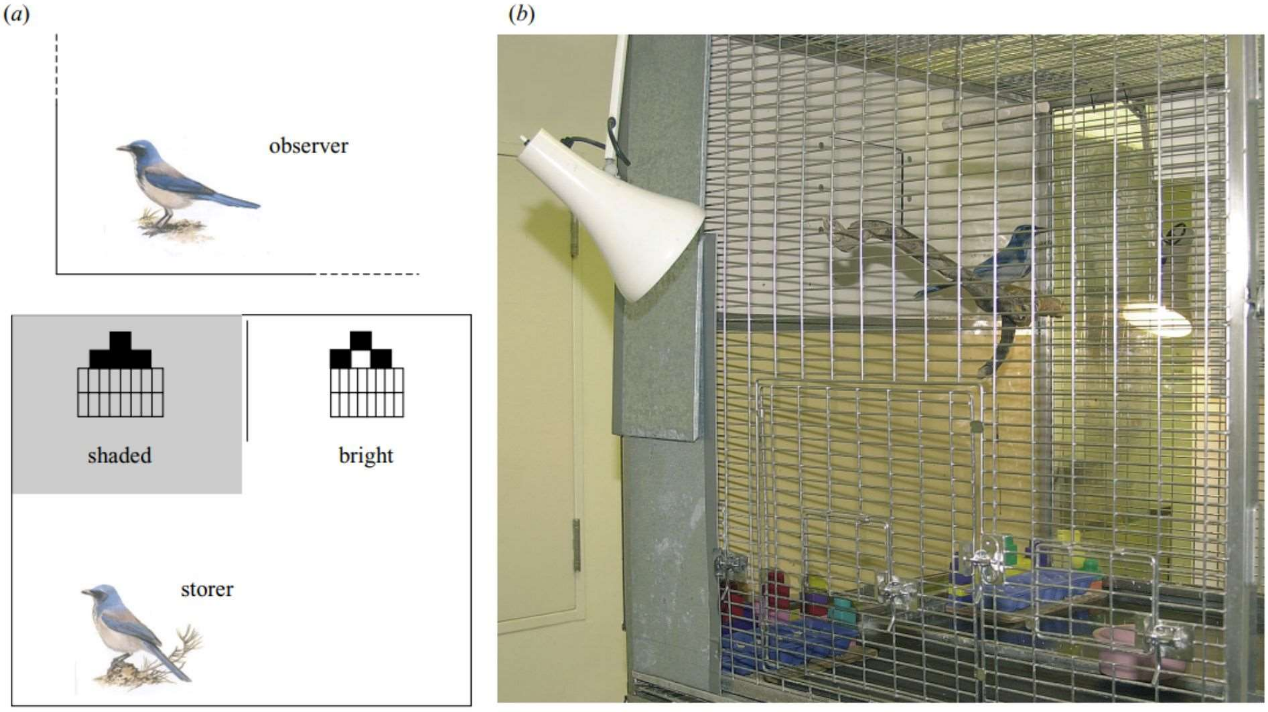


Figure 4 shows the experimental setup, which consisted of two caching areas: one illuminated and one shaded. The areas were located in a compartment for scrub jays, with another cage for the observer. Each area had a caching tray and a lamp placed 54 cm above the ground. However, only the lamp on the illuminated area was lit, creating a difference in illumination and temperature between the two areas. This figure is adapted from Dally et al. 2004.

Here, a complementary behaviour-reading hypothesis of the jays' behaviour can be considered alongside Dally and colleagues' analysis. It is possible that the jays have learned through experience that hiding food in certain areas, such as shaded or less visible areas, reduces the risk of theft, without necessarily understanding the perceptions or intentions of other birds. Scrub jays may use simple rules to hide food when in the presence of a conspecific, rather than relying on a detailed understanding of the observer's mental state. This behaviour likely evolved because it provides an adaptive advantage, increasing the likelihood of conserving one's own food supply, without requiring the animals to understand the cognitive processes of others. Furthermore, the results of the experiment can be interpreted as behaviour-reading rather than mind by comparing it with the chimpanzee experiment and the direct line mechanism between a subject's eyes and the object. In the chimpanzee experiment, a visual barrier between the individual and the food acted as an obstacle that the chimpanzees could interpret directly. If the barrier was transparent, the line from the conspecific to the food was clear. If it was opaque, the trajectory was blocked. In the experiment with scrub jays, the discriminating factor was the amount of light rather than the presence of a physical barrier. Just as there are varying degrees of opacity and transparency, there are also varying levels of light intensity. A well-lit area can be compared to a transparent barrier, while a shaded area can be compared to an opaquer barrier. The hiding behaviour of scrub jays in shaded areas when in the presence of conspecifics may not necessarily indicate an attribution of visual ability to conspecifics. Rather, it may simply reflect an assessment of environmental conditions, as shade can obstruct the trajectory between the conspecific and the tray, increasing the likelihood of food remaining hidden.

The third experiment worth mentioning, again with scrub jays, is that of Stulp, Emery, Verhulst, and Clayton (2009). In this experiment, scrub jays always had to hide food, but there were three conditions under which they had to do so: (i) in the private condition, a scrub jay hid food in the absence of observers, just as in the previous experiments. In this scenario, the bird had no reason to worry about theft since no one else knew the location of the food. (ii) In the public condition, a conspecific could see and hear the scrub jay while it was hiding the food. This situation simulates a scenario in which the scrub jay hiding the food must consider the possibility of theft, as a potential thief knows the location of the stock. (iii) In the third condition, the conspecific could only hear but not see the scrub jay hiding the food. To make the experiment more revealing, the researchers used two types of substrates in the trays where the scrub jays could hide food: one with stones that produced noise when moved and the other with sand, which offered a quieter option for hiding food. The results showed that the scrub jays adapted their caching behaviour based on the situation. In the third condition, the scrub jays tended to choose the tray with sand, indicating a preference for a quieter strategy for hiding food. In the other conditions, however, the scrub jays showed no preference for which tray to use to hide the food. This pattern is consistent with the interpretation mentioned earlier. The researchers used the mind-reading interpretation to explain the behaviour of the scrub jays, believing that they were able to attribute hearing or vision to their conspecifics. This is why it would prefer a less noisy bowl in the third condition. However, it is important to note that a complementary behaviour-reading hypothesis is also possible. Similar to Dally et al.'s (2004) experiment on vision, jays may have learned over time that sounds must be unimpeded to have an effect on behaviour. The sound volume varies according to the light gradations in the first experiment with scrub jays. Stulp, Emery, Verhulst, and Clayton's (2009) experiment demonstrates that even when we use other sensory channels besides sight to design experimental protocols, we encounter the same logical problem that cannot be solved. The jay may rely on another individual's presence and surroundings instead of assuming it has the cognitive ability to hear.

4.1.2 Knowledge attribution experiments

In 2001, Hare and colleagues published the results of three competitive experiments with chimpanzees. These experiments aimed to explore the ability of subordinate chimpanzees to understand the knowledge or ignorance status of dominant chimpanzees regarding the location of hidden food. All three experiments had variations but followed the same set-up. In three separate experiments, a subordinate chimpanzee competed with a dominant chimpanzee for hidden food behind two opaque barriers. The barriers were positioned in the central area that connected the rooms of the two chimpanzees (see Figure 5). Prior to the competition, the subordinate chimpanzee could observe the dominant chimpanzee from his room as the experimenter hid the food, or he could infer that the door to the rival's room was closed, preventing the dominant chimpanzee from witnessing the hiding of the food. The study discovered that subordinate chimpanzees were less likely to attempt to take food from a hiding place during competition if they had previously observed the dominant chimpanzee witness the hiding of food, compared to when they had not witnessed such an event. The results suggest that subordinate chimpanzees were able to attribute knowledge or ignorance to their dominant rivals based on whether they had seen the food hiding or not.

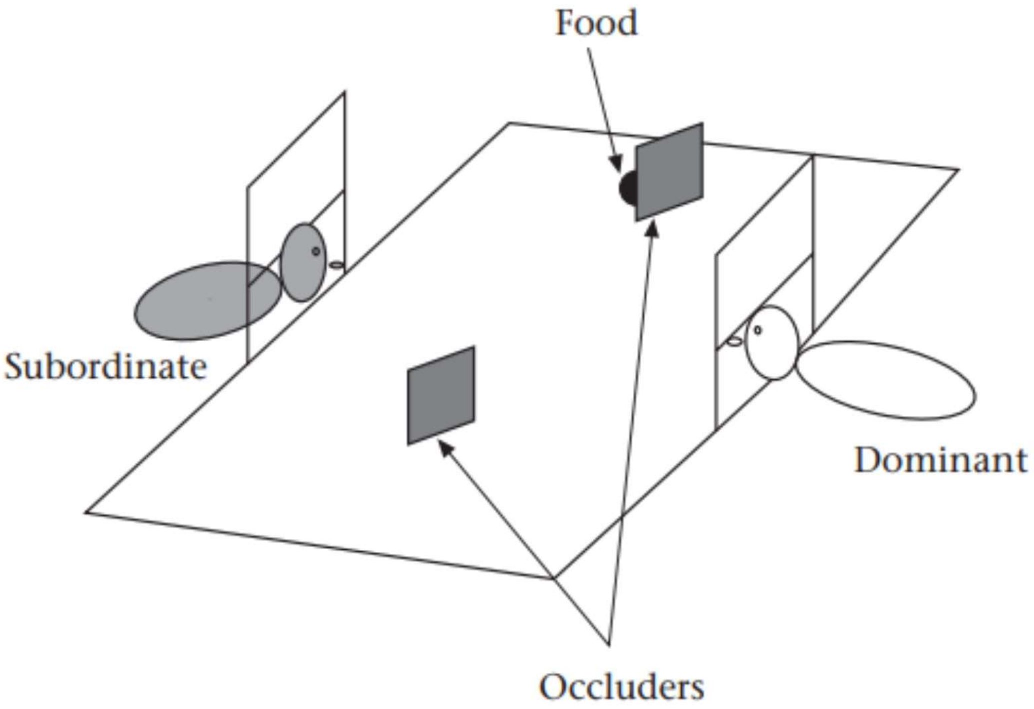


Figure 5 depicts the general setup of the proposed chimpanzee competition experiment designed to test their ability to attribute knowledge to others. Adapted from Hare et al., 2001.

Although this experiment aims to investigate the ability of a chimpanzee to attribute knowledge states to one of its conspecifics, the experiment is based on the same mechanism as the first experiment mentioned by Hare et al. (2004). Attributions of knowledge or ignorance in this context are based on whether or not the hiding of food was seen. The subordinate chimpanzee can only attribute knowledge or ignorance to the dominant if it can first attribute the ability to see. This ability can always be explained by the direct line of sight between the chimpanzee's eyes and the food. Therefore, as in the first experiment, the subordinate chimpanzee can only recognize and respond to observable patterns of behaviour, reading the dominant's behaviour and not its mind. Once again, a complementary behaviour-reading hypothesis is possible.

In addition, other experiments, such as those conducted by Kuroshima and colleagues (2002, 2003) or by Gòmez and Texidor (1992), can be explained by the same complementary behaviour-reading hypothesis described earlier.

4.1.3 False belief attribution experiments

Lurz (2011) cites an experiment that tests false belief attribution, which was conducted by Kaminski and colleagues in 2008. The experiment involved two chimpanzees, a subject, and a contestant. A table with three cups was placed in front of the chimpanzees, separated from them by plexiglas panels (Figure 6). Two types of rewards are presented in this experiment: a high-quality reward (e.g., a piece of banana) and a low-quality reward (e.g., a piece of apple). The high-quality reward is always hidden under one of the cups on the main table, while the low-quality reward is placed on an adjacent table that is only accessible to the subject. In each trial, the experimenter displays and hides the low-quality reward next to the subject as a 'safe' option, and then hides the high-quality reward under one of the cups on the main table. The experimenter moves the cups, either lifting and repositioning them with the reward in the same place (lift) or shifting them with the reward in a different place (shift). Depending on the experimental condition, these movements may be visible to both participants or only to the subject. If the movements are only visible to the subject, it is because an opaque barrier is placed between the competing chimpanzee and the table with the cups. There are four experimental variables: the Known Lift, the Known Shift, the Unknown Lift, and the Unknown Shift. Lurz (2011) cites an experiment that tests false belief attribution, which was conducted by Kaminski and colleagues in 2008. The experiment involved two chimpanzees, a subject, and a contestant. A table with three cups was placed in front of the chimpanzees, separated from them by plexiglas panels (Figure 6). Two types of rewards are presented in this experiment: a high-quality reward (e.g., a piece of banana) and a low-quality reward (e.g., a piece of apple). The high-quality reward is always hidden under one of the cups on the main table, while the low-quality reward is placed on an adjacent table that is only accessible to the subject. In each trial, the experimenter displays and hides the low-quality reward next to the subject as a 'safe' option, and then hides the high-quality reward under one of the cups on the main table. The experimenter moves the cups, either lifting and repositioning them with the reward in the same place (lift) or shifting them with the reward in a different place (shift). Depending on the experimental condition, these movements may be visible to both participants or only to the subject. If the movements are only visible to the subject, it is because an opaque barrier is placed between the competing chimpanzee and the table with the cups. There are four experimental variables: the Known Lift, the Known Shift, the Unknown Lift, and the Unknown Shift. The study authors hypothesized that if the subject chimpanzees were able to attribute beliefs, they would understand that in cases of "unknown shift" their competitor would likely choose the wrong cup because of his false belief about the location of the food. Therefore, in these situations, the subject chimpanzee would be expected to choose the cup on the table, which is more likely to contain the food. In unknown-lift situations, the subject chimpanzees should have expected the competitor to correctly choose the cup with the food, thus choosing the safe but less desirable reward in their cubicle. However, the results showed no significant difference in the chimpanzees' choices between the 'unknown-shift' and 'unknown-lift' trials. The authors interpreted this behaviour as a possible indication that chimpanzees may not be able to attribute beliefs. This is consistent with previous research conducted by Call and Tomasello in 1999.

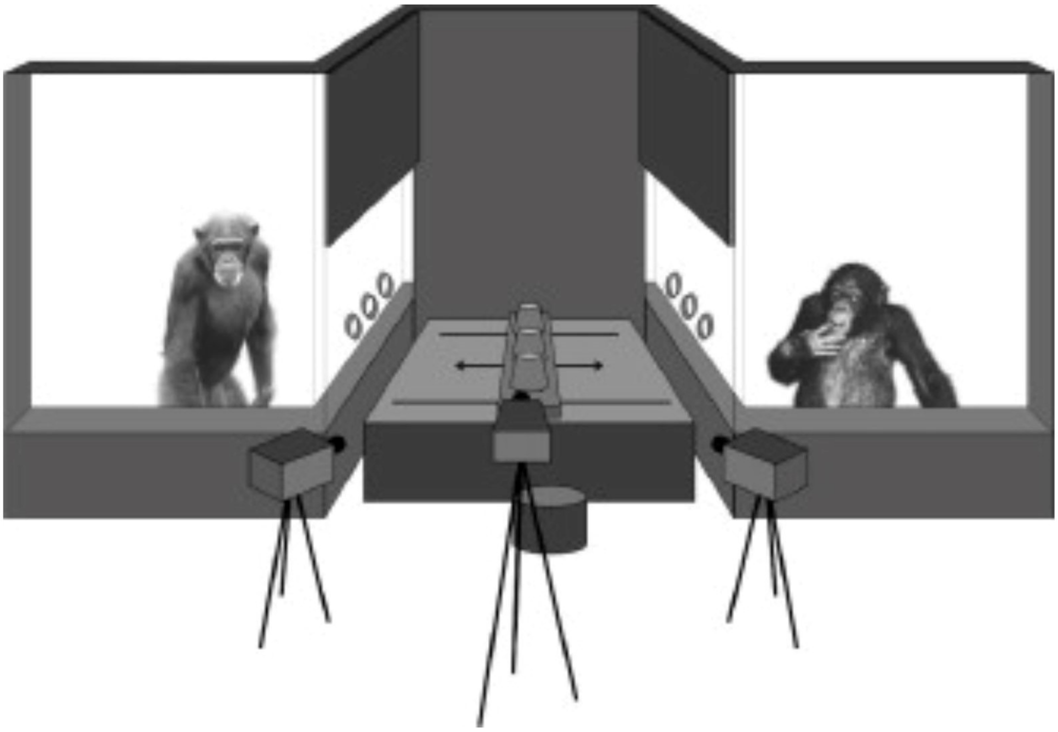


Figure 6 shows the experimental apparatus used to test false beliefs in chimpanzees, reprinted from Kaminski et al. (2008).

The interpretation of the experiment's negative results with chimpanzees conducted by Kaminski and colleagues suggests that there could be multiple reasons why the chimpanzees did not demonstrate evidence of understanding the competitor's false beliefs, beyond their ability to attribute beliefs (Lurz 2011). If the chimpanzees did not observe the contestant's choice behaviour, they may not have comprehended that it was the contestant's turn to choose first. The chimpanzees' behaviour may have been influenced by their perception that it was their turn, rather than their inability to attribute beliefs. Additionally, they may have believed that one of the containers on the table still contained higher quality food, despite what they observed during the food search phase. This could explain why they chose one of the containers on the table instead of the low-quality 'safe' option. This experiment is valuable because it demonstrates that a negative result does not necessarily rule out the cognitive ability being investigated in the animal subject.

It is also worth mentioning the following experiment conducted by Tschudin in 2006 on dolphins. This experiment, unlike the one mentioned above, claims to have yielded positive results regarding the attribution of false beliefs in nonhuman animals. Four dolphins were trained to respond to a specific gesture, tapping, performed by an experimenter to indicate the presence of a reward (a fish) in one of two boxes. During training, the dolphins observed the experimenter watching while a second experimenter hid the fish under one of the two boxes. The dolphins were unable to see which box was chosen. The experiment began with the first experimenter approaching the boxes and touching the one containing the food. The dolphins were then given the opportunity to choose a box. If they selected the box with the food, they received the fish. The subsequent phase of the experiment involved a series of tests that depended on the first experimenter's accurate or inaccurate beliefs. In the false belief test, the first experimenter witnessed the second experimenter concealing the fish. In the experiment, the first experimenter walks away and the second experimenter swaps the position of the two boxes. When the first experimenter returns, they touch the wrong box, mistakenly believing it to be the correct one. The dolphins are then asked to choose a box. This scenario is repeated in the true belief test. The only difference between the two tests is that in the second test, the first experimenter observes the exchange of position between the two boxes. In the false belief tests, it was expected that if dolphins can attribute beliefs, they would understand that the first experimenter has an incorrect belief and choose the box that was not touched. Similarly, in true belief tests, they were expected to follow the gesture of the first investigator and choose the box that was touched. The results indicate that dolphins can attribute beliefs by choosing appropriately in the false and true belief tests, supporting the mind-reading hypothesis. However, it is important to note that their behaviour could also be explained by a complementary behaviour-reading hypothesis, as suggested by Lurz (2011). For instance, dolphins may have learned from experience that certain behaviours or situations are associated with an increased likelihood of food being in a particular location, such as the absence of the experimenter during the exchange of boxes. This type of learning does not require attributing a false belief to the experimenter, but rather a form of conditioning or reading of clues based on past experience. Furthermore, the potential existence of unintentional signals from the experimenter is another alternative explanation. Animals may develop a remarkable ability to perceive and react to subtle, involuntary signals given off by humans, such as slight changes in body language, posture, or facial expression. These cues could indicate the correct location of food without requiring an understanding of the experimenter's beliefs.[64]

Upon analyzing the studies highlighted for each of the aforementioned categories (perceptual abilities, knowledge, false beliefs), two common features can be identified as the basis for the inability to solve the logic problem: (i) when the animal is called upon to predict the behaviour of another individual, whether it does so by behaviour-reading hypothesis or mind-reading hypothesis, it is still relying on the same observed behaviours and environmental contexts; (ii) animals may rely on strategies based on learning and past experiences to solve tasks. According to current knowledge, experimental protocols should eliminate these two features to minimize or exclude the possibility of animals relying on behaviour-reading strategies.

Lurz (2011) proposes a theoretical basis to support the validity of his experimental protocols in eliminating the two conditions. Before delving into the details of these protocols, it is important to clarify the theoretical basis. Lurz calls this theory the Appearance-Reality Mindreading Theory. The theory of Appearance-Reality Mindreading (ARM) has its origins in earlier studies by scholars such as Humphrey (1976) and Gallup (1982). Later work by Krachun (2008), Krachun et al. (2010), and Lurz (2011b) further deepened the theory. According to ARM theory, the ability to attribute to others not only the ability to perceive objects but also to perceive them differently than how they appear may have been a significant evolutionary advantage. This ability would have enabled animals to improve their survival and reproduction strategies, as well as to explore environments characterized by perceptual illusions more effectively. To clarify this concept, Lurz offers the example of a

---

[64] In addition to the experiments conducted by Tschudin and Kaminski, other experiments have been conducted by Krupeneye et al. (2016), Buttelmann et al. (2017), and Hayash et al. (2020) which have investigated the attribution of false beliefs in non-human animals. Nevertheless, these studies also fail to address the logical problem. Their findings can be interpreted in accordance with either the mindreading hypothesis or its complementary behaviour-reading hypothesis.

chimpanzee observing a conspecific looking at an appetizing fruit in a forest so dense that the lighting conditions make the same fruit seem unappetizing or unrecognizable. If the first chimpanzee is able to attribute to the second chimpanzee the perceptual appearance state of the fruit, it might exploit this knowledge to grab the fruit. The ARM theory suggests that in illusory contexts, where there is a marked discrepancy between perceptual appearance and objective reality, animals capable of attributing perceptual appearance states have an advantage in predicting the behaviour of naive agents who rely on the appearance of a certain state of affairs rather than its actual characteristics. Another example is a predator using camouflage to deceive its prey. An animal with the ability to recognize perceptual appearance states could identify that the prey is fooled by the predator's appearance, allowing it to avoid making the same mistake. According to ARM Theory, this type of mental ability would offer significant adaptive advantages, facilitating prediction and potentially manipulation of others' behaviour in complex situations. The theory is in line with the Machiavellian intelligence hypothesis[65], which posits that the main advantage of mind reading is to enhance the predictability of the behaviour of others, particularly those of the same species. The ARM Theory states that illusory environments are important contexts for attributing perceptual appearance states, which allows animals to predict the behaviour of others in situations where those who rely solely on objective facts cannot. The integration of the ARM theory with the concept of Machiavellian intelligence offers a more comprehensive perspective on how the ability to attribute perceptual appearance states and distinguish between appearance and reality may have evolved as a crucial adaptation for navigating illusory environments. This enhances the ability of animals to predict and manipulate the behaviour of others, increasing their chances of survival and reproductive success[66]. Lurz proposes two protocols that exploit

---

[65] Machiavellian intelligence is the ability to understand and manipulate social behaviour through cunning and strategy. The term is derived from Niccolo Machiavelli, a Renaissance philosopher known for his writings on the use of power and manipulation in politics. In the field of evolutionary psychology and ethology, Machiavellian intelligence is a theory that suggests certain species, especially primates, have developed intricate social and cognitive abilities to navigate the competitive and cooperative social dynamics of their group. The theory also suggests that some species have the ability to deceive others. These abilities include recognizing deception, forming alliances, understanding dominance relationships, and using sophisticated social tactics to gain personal or group advantage. The hypothesis of Machiavellian intelligence suggests that advanced cognitive skills are naturally selected because they offer significant survival and reproductive advantages in complex social contexts.

[66] Lurz's ARM theory posits that individuals can understand that others may perceive a state of affairs differently from how it appears. This aligns with multimodal mind theory, as this understanding requires processing a larger

the principle of illusory contexts to overcome the logic problem for attributing perceptions and beliefs.

4.2 How to solve the logic problem for perception

Considering the characteristics of the logic problem, the design of existing experimental paradigms, and ARM theory, Lurz (2011) proposes a modified version of the EP protocol. The aim is to make the mind-reading hypothesis more likely and to reduce the likelihood of the complementary behaviour-reading hypothesis. The experiment aims to create an illusory environment that makes it impossible for an individual to pass the test without ToM, since his or her perceptions of a given state of affairs will necessarily be different from those of other subjects. In order to pass the test, a subject must be able to represent what another animal is perceiving and acknowledge that the way they perceive an object may differ from the way the other animal perceives it. Lurz proposed an experimental procedure consisting of three steps. The first step is to create an illusion by exposing an animal to an environment in which objects appear different from their actual form. This is done to create an illusion in which the animal perceives objects as having property F, even though they actually possess property G. During the second stage, the animal and any other potential observers are ensured to not witness any behaviour associated with the recognition of objects as G. Instead, they only observe object-related behaviours when they are perceived as F. This establishes a precedent in the animal's learning experience that associates certain behaviours with perceived property F, not actual property G. The final stage tests whether the animal can anticipate that another agent, whether it be another animal or an experimental structure that mimics the behaviour of an

---

volume of information than simply attributing perceptions to others. In the former case, an individual must consider both the actual reality of the state of affairs and another individual's perception of that state of affairs. This requires taking into account how things actually are and how they may appear to another subject, who may perceive them differently due to perceptual illusions, different angles, past experiences, etc. In the second case, the individual only needs to handle information related to direct perception of the state of affairs, without having to consider any discrepancies between appearance and reality.

agent, will act in a certain way when confronted with an object that is actually G but appears to be F due to the illusory setting.

To clarify how to put these three steps into practice, Lurz proposes different examples of EP protocol modification. These include the use of transparent colored barriers, size-distorting barriers, and deceptive amodal completion stimuli[67]. Let us review them briefly.

4.2.1 ARM transparent colored barriers experiment

Based on the previous study by Hare and colleagues (2000), this experiment involves two chimpanzees, one with a dominant status and the other subordinate, being placed in an environment to test their understanding of each other's visual perspective and subsequent actions. The set-up includes two separate rooms for the chimpanzees, flanked by a central area where the actual testing takes place. The central area hosts a competition between two types of bananas: a real, luscious yellow banana and a fake, unappetizing orange banana made of plastic (see figure 7). Colored transparent barriers are introduced, altering the appearance of the objects behind them. A transparent barrier does not affect the perceived color of objects, while a red barrier alters color perception. For instance, yellow objects appear orange and orange objects appear darker. Prior to the main experiment, the chimpanzees undergo a preliminary phase in which they learn the rules of competition. They learn that they can only take bananas that are within reach and that the dominant chimpanzee tends to ignore the orange bananas after discovering that they are not real food. This stage establishes the foundation for chimpanzees to form expectations based on their experiences. A crucial retraining phase follows for the subordinate chimpanzee, during which it is exposed to the colored barriers and learns how they affect the appearance of the objects behind them. This phase is critical because it teaches the subordinate the illusion created by the red barrier, laying the groundwork for the next test. Then comes the test phase. A yellow banana was placed behind the red barrier and an orange banana behind the clear barrier. The

---

[67] Amodal completion is a perceptual phenomenon in which an observer's cognitive system infers the presence of a part of an object or sound that is not directly perceptible through the senses due to occlusion. In other words, it is the mental process that allows us to 'complete' an object partially hidden from view or an interrupted sound and perceive it as a whole. Despite lacking direct sensory information about the occluded part, our brain integrates available perceptual data to form a complete representation of the object or sound.

doors were opened slightly for the chimpanzees to observe the setup and each other. The subordinate was given a slight advantage to make its choice. If the subordinate chimpanzee understands that the dominant chimpanzee sees the yellow banana as orange due to the red barrier and will not choose it, it may choose the yellow banana for itself, demonstrating a form of ToM ability. On the other hand, if the subordinate bases their decision solely on past observations without understanding the visual illusion experienced by the dominant, they may make a different choice, revealing a more behaviour-reading reasoning process.

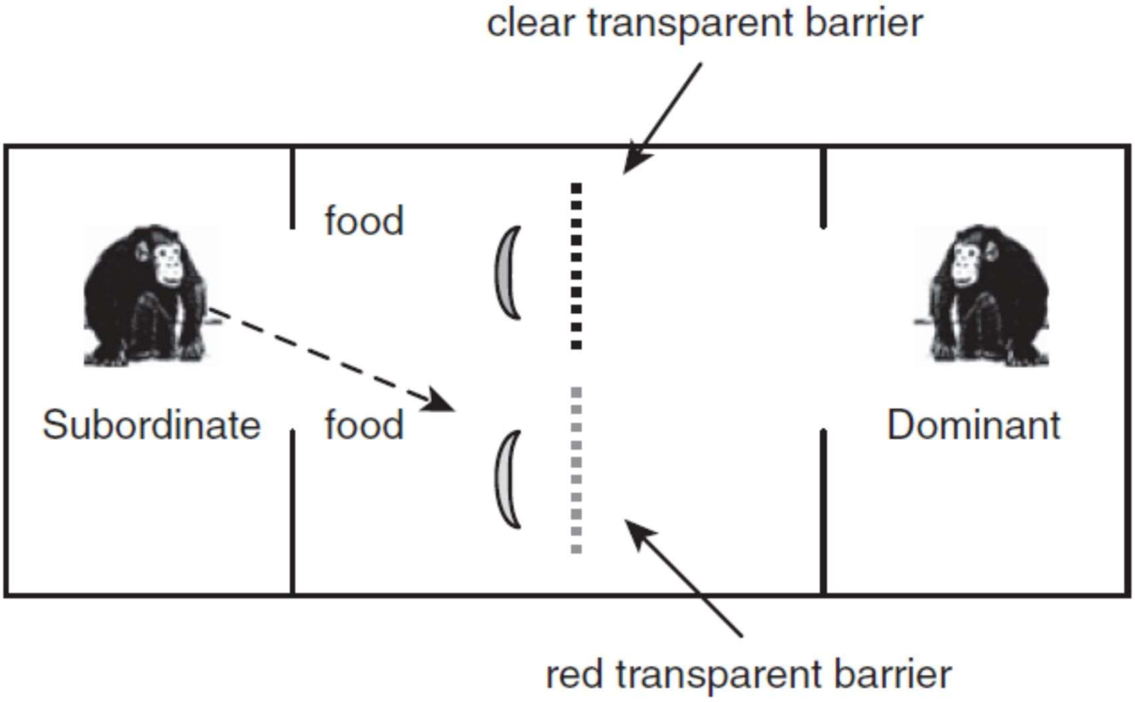


Figure 7 shows the expected performance of a mind-reader chimpanzee in the colored barrier test, as reprinted from Lurz (2011).

It is not possible to interpret the results of this experimental paradigm according to a complementary behaviour-reading hypothesis because the dominated chimpanzee would act differently than in the mind-reading hypothesis. The data on which the dominated chimpanzee relies to predict the conspecific's behaviour according to the two hypotheses are no longer the same due to the illusory context. If a chimpanzee in a subordinate position has ToM, it is able to process the information from the ability to understand that the banana behind the red barrier has different properties for the conspecific than it does in reality. This means that the subordinate chimpanzee can understand a state of affairs that it cannot perceive directly. From its point of view, the banana is yellow and there is no way for it to perceive it as orange, as the dominant chimpanzee is doing. In cases where the dominated chimpanzee is simply reading the dominant's behaviour, the dominant will perceive the banana as yellow, even if the banana is behind the red barrier. This is in contrast to the chimpanzee's perception of the banana as orange. By incorporating the illusory setting, Lurz (2011) successfully created an experimental paradigm that does not rely on the same observed behaviours and environmental contexts to test the two hypotheses (behaviour-reading and mind-reading).

4.2.2 ARM size-distorting barriers experiment

The second experiment is based on previous research by Carla Krachun and colleagues (2009). They used magnification and minimization lenses to test the ability of chimpanzees to discern the true size of objects despite visual distortions. In Krachun's study, chimpanzees were shown two grapes behind different lenses. One grape appeared larger through a magnification lens and the other smaller through a minimization lens. Initially, the chimpanzees chose the grape that appeared larger behind the magnifying glass, indicating their susceptibility to visual distortion. However, upon being presented with the choice again, over half of the chimpanzees selected the grape behind the minimization lens, demonstrating their ability to perceive reality beyond the visual distortion. This finding allows for two possible interpretations. The first is that the chimpanzees comprehended the disparity between the grape's appearance and its true size. The image-tracking hypothesis suggests that the chimpanzees associated the distorted image with the real object behind the lens, without fully understanding the concept of appearance versus reality. Building on this foundation, Lurz proposed an experiment involving a subordinate and a dominant chimpanzee competing for bananas of different sizes. The bananas were placed behind barriers that distorted the size of the bananas or left their appearance unchanged. They learn which bananas are safe to retrieve based on the dominant's line of sight and choices. The experiment involves testing whether subordinate chimpanzees can use their understanding of the dominant's perspective to make strategic choices about which bananas to take. Barriers are introduced to change the apparent size of the bananas, and the subordinate chimpanzees must use their knowledge to make the right choices (see figure 8). For instance, in a particular scenario, a small banana placed behind a magnification barrier may seem larger to the dominant individual, which could discourage the subordinate from selecting it due to the anticipation of competition. On the other hand, a large banana placed behind a minimization barrier may appear smaller, making it a safer option for the subordinate if it assumes that the dominant will overlook it. The experiment aims to determine whether the subordinate chimpanzee is making decisions based on a true understanding of the dominant's perspective (mind-reading) or simply reacting to observed behaviours without catching the underlying visual deception (behaviour-reading) by the same mechanism as in the colored barrier experiment.

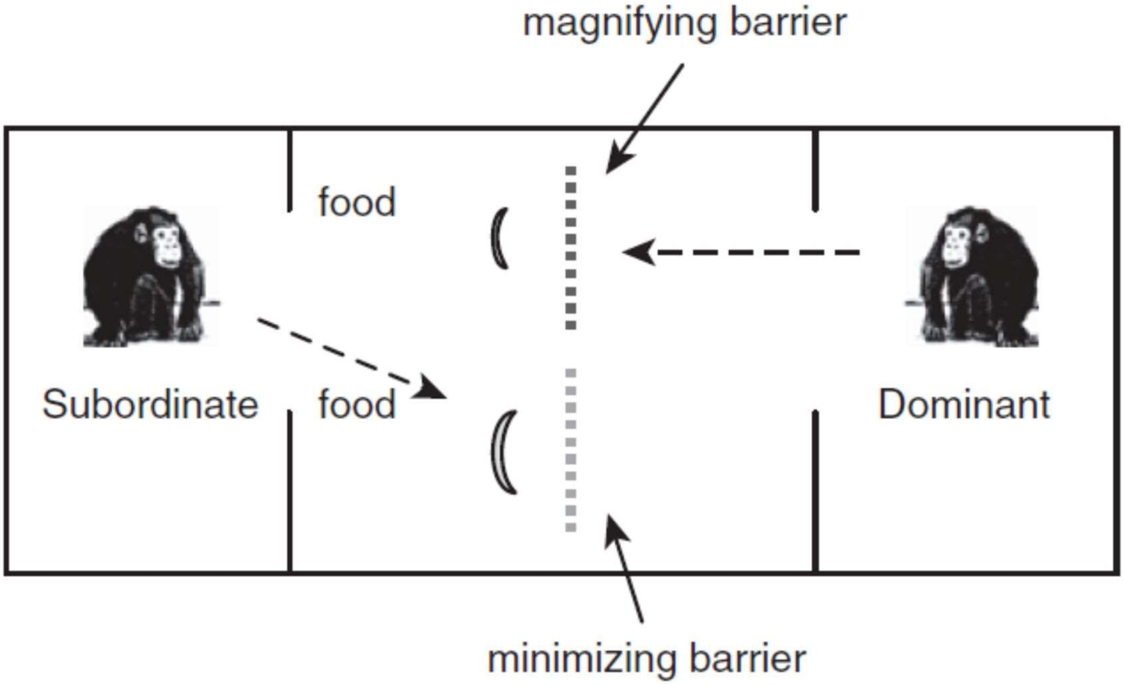


Figure 8 shows the expected performance of a mind-reader subordinate in the magnifying/minimizing test, as presented in Lurz's (2011) work.

4.2.3 ARM deceptive amodal completion stimuli experiment

The third experiment investigates whether chimpanzees can anticipate a human's actions based on their understanding of how a visual stimulus, an optical illusion of a partially hidden object, was previously perceived by the human. The protocols used to study the attribution of mental states in infants (Southgate et al. 2007; Song & Baillargeon 2008) serve as the basis for this experiment. Before the chimpanzee enters the test area, an experimenter sets up a table in front of the chimpanzee's cage. On the table, there is a box with one side painted to look like an incomplete triangle. The box is placed behind an object that partially hides it, creating the illusion of a complete triangle. Additionally, an attention-grabbing object, such as an apple or a noisy toy, is placed on the table (see figure 9). During the initial phase, known as habituation, the chimpanzee observes the table and its objects directly. Another experimenter stands in front of the table, with the same perspective as the chimpanzee but facing away from it. This experimenter then interacts with the object of interest, such as examining or biting the apple, before placing it inside the box and closing the lid. During the testing phase, the experimenter who interacted with the object turns or walks away, losing sight of the table. At this point, the other person moves the box, revealing that the triangle design is actually 'amputated' and not as complete as it appeared. Subsequently, a second box with a complete painted triangle is introduced and placed at the other end of the table. The objective of the experiment is to determine whether the chimpanzee will choose the box with the complete triangle when the experimenter returns and reaches for one of the two boxes. This is based on the chimpanzee's previous observation that it appeared that way to the experimenter. If the chimpanzee displays surprise when the experimenter selects the box with the amputated triangle, or if it tends to gaze at the box with the complete triangle before the experimenter acts, this could suggest that the chimpanzee is capable of attributing a mental state to the human. Specifically, the chimpanzee may understand that the human might believe that the object is in the box with the triangle that appears complete.

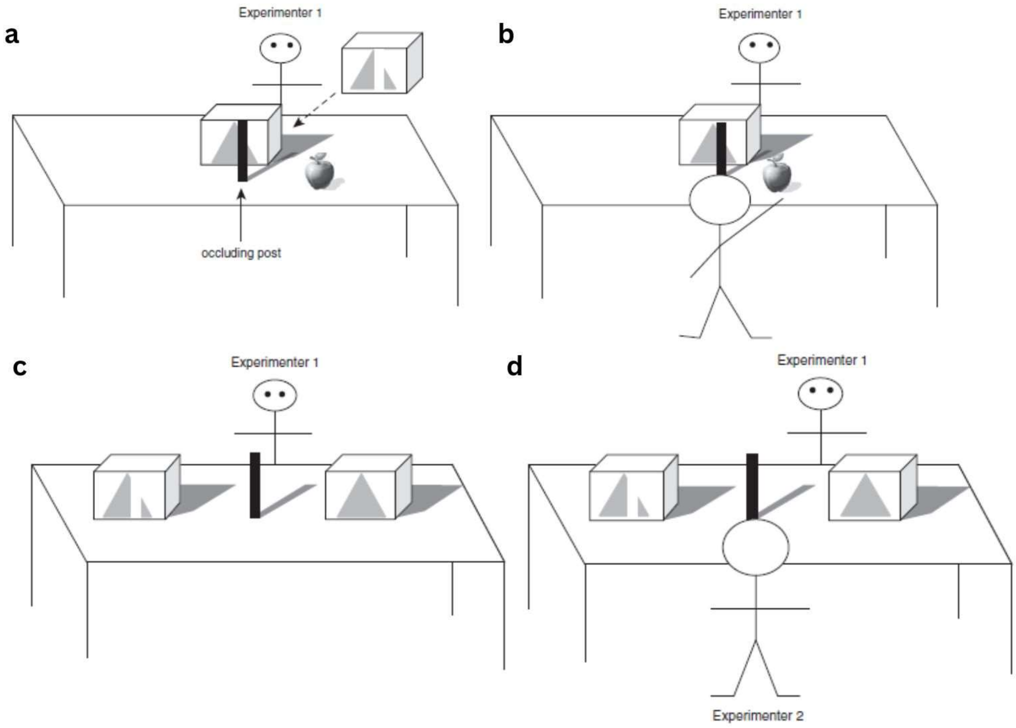


Figure 9 shows (a) the initial setup of the table, (b) the habituation phase (c) Experimenter 1 revealing the deceptive amputated-triangle box and placing the whole triangle box on the table, and (d) the test trial where Experimenter 2 returns and reaches for one of the two boxes. This figure is reprinted from Lurz (2011).

The main idea is that if a chimpanzee relies only on objective facts, and acts according to a complementary behaviour reading hypothesis during the habituation phase, such as the position of the apple and the experimenter's view, one would logically expect the chimpanzee to anticipate that the experimenter would choose the box associated with the apple, i.e., the one with the amputated triangle. From the chimpanzee's perspective, there is no reason to believe that the experimenter would prefer a different box, since the last known location of the apple was in the box with the amputated triangle. However, if the chimpanzee displays surprise or confusion when the experimenter selects the box with the complete triangle, it would suggest that the chimpanzee had a different expectation, indicating a deeper understanding that includes the ability to attribute visual perceptions or mental states to the experimenter. In other words, the chimpanzee might understand that despite its direct knowledge of the apple's location, the experimenter might have had a different visual experience that led them to believe that the apple was in another box, thus acting according to mind-reading hypotheses.

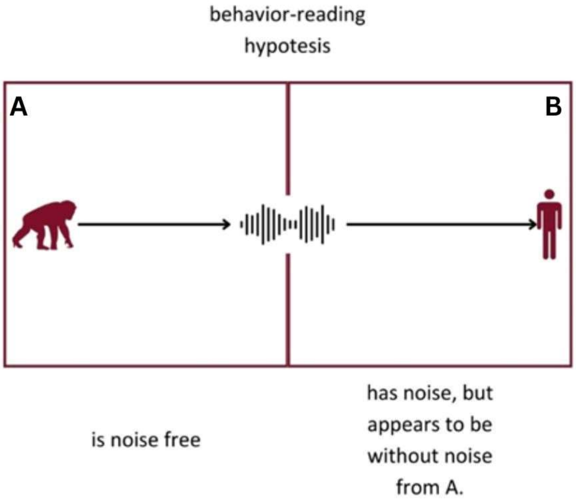
Upon examining the three experimental protocols presented, it becomes clear that none of them exhibit the two features that are necessary for solving the logic problem. The first feature, which we have already pointed out, is that the data on which the subjects rely differ due to the illusory context, whether they act by mind-reading hypothesis or by complementary behaviour-reading hypothesis. In all three experiments, the chimpanzees were placed in a novel condition where they had to predict another subject's behaviour. This task was also novel, meaning they could not rely on past experience or learning strategies to solve it.

It is worth noting that this approach has been challenged, particularly by Halina (2015). Halina challenges Lurz's approach to mind-reading theory in animals by invoking the theoretician's dilemma identified by Hempel (1958), which suggests that if the theoretical terms of a scientific theory establish unambiguous connections between observable phenomena, then these theoretical terms can be eliminated and replaced by laws that directly connect observable antecedents and consequents. Halina applies the principle of objectivity to the problem of mind-reading. She argues that any experimental success in attributing mental states to animals can be traced back to observable behavioural regularities, making the attribution of mental states unnecessary. In other words, if one animal successfully predicts another's behaviour based on inferred mental states, it could just as easily rely on observable cues, obviating the need to attribute complex mental states. Halina suggests that convincing evidence for ToM can only be provided by a series of tests that demonstrate an animal's ability to consistently respond to a variety of situations indicative of different mental states (labeled S, S*, S**). This approach is based on Mill's (1843) method of agreement, which aims to identify the common cause of an observed effect across different situations. However, Lurz may argue that these different situations could be viewed as variations of a single category of stimuli (S), and therefore test results could still be explained through a complementary behaviour-reading hypothesis, inferring a direct relationship between observable stimuli and resulting behaviours (S -> B). Both Halina and Lurz agree that a single test is insufficient to demonstrate ToM in non-linguistic animals. Halina emphasizes that only a series of tests demonstrating consistency in animal behaviour across different situations could provide convincing evidence of mind reading, rather than simple behavioural reading. Lurz also recognizes the importance of a series of tests but points out Halina's error in assuming that it is not necessary to develop individual tests capable of discriminating between complementary behavioural reading hypothesis and the mind-reading one. Lurz suggests that it is important to design experiments that isolate mental attribution abilities, although no single test can be definitive. This is a reasonable and logical approach to take.

4.2.4 Multimodal-shift experience-projection protocol

At this point, an adaptation of Lurz's experimental procedure is proposed, which is capable of exploitation of the mechanism of multimodal shift. The purpose of this approach is twofold: (i) It expands the range of experimental paradigms available for testing ToM in non-linguistic animals. (ii) On the other hand, by exploiting the multimodal shift, I believe that simpler tests can demonstrate the presence of ToM in animals even in the absence of illusory settings. First, let us discuss how to utilize the multimodal shift in the same setting proposed by Lurz. To provide a practical indication of how such an experiment might be conducted, we can build on the results of several previous studies conducted on captive chimpanzees (Pan troglodytes) (Call & Tomasello, 1994; Tomasello et al., 1994; Hostetter et al., 2001, 2006; Leavens et al., 1998, 2004a, 2004b; Russell et al., 2005) that performed a multimodal shift based on the orientation or visual regard of other test subjects, within an interspecies communicative context and in some cases with conspecifics. The hypothetical experiment is as follows.

The experiment area is divided into two zones: an area labeled A (where the animal is located) and an area labeled B (where the keeper is located). Communication in both areas can be disrupted by playing a disturbing sound. Learning Phase: In this phase of the experiment, a chimpanzee learns in a cooperative setting that it can obtain food from the experimenter by attracting his/her attention with a signal. If the area is disturbed by a noise the chimpanzee will have to use a visual cue to attract the experimenter's attention. If the area is not disturbed by a noise the chimpanzee will have to use a vocalization to attract the experimenter's attention. The chimpanzee is considered to learn how to obtain food when it consistently uses visual or auditory signals based on environmental conditions without making mistakes. When, in the presence of noise the chimpanzee changes communicative mode it makes a multimodal shift. It must be ensured during the learning phase that the chimpanzee does not exhibit a multimodal shift in the absence of noise but performs a multimodal shift when noise is present. That is, the chimpanzee must exhibit behavioural consistency. The exploration phase follows. During the exploration phase the chimpanzee is made aware of a new configuration of the experimental area. During this phase the experimenter leaves the experimental area and does not experience this new configuration. In the new configuration, area A is silent. The chimpanzee cannot hear any disturbing sound while it is in area A. area B, on the other hand, is strongly sound disturbed. The chimpanzee while in area B can hear the disturbance sound. The disturbance sound audible in area B cannot be heard in area A. This is followed by the last phase of the experiment, the anticipation test phase. In the anticipation test phase, the experimenter re-enters the experimental area by positioning himself in area B (heavily sound disturbed). The chimpanzee must at this point attract the experimenter's attention to obtain food. In this phase, the chimpanzee is asked to interact freely for a set period of time (one or more minutes) with the researcher. At the end of the set time, the researcher gives the chimpanzee food. The study observes whether the chimpanzee, which cannot hear the noise present in area B, being in area A, is able to understand that the experimenter cannot hear any vocalizations because of the disturbance sound. If the chimpanzee has the ToM ability, it will perform a multimodal shift to attract the experimenter's attention. That is, it will use visual signals to attract its attention. On the other hand, if it does not have the ToM ability it will use auditory signals by not hearing any disturbing sounds. Control condition: In the control condition, the experiment is performed as just described, but the exploration phase is eliminated. In this case, the chimpanzee, although having ToM, not having experienced the disturbance sound present in area B, should not perform a multimodal shift. Consequently, the chimpanzee, hearing no sound, in the control condition during the anticipation test phase will always try to attract the experimenter's attention with a sound and never with a visual signal. It is worth noting that in the experimental condition during the anticipation test phase, the multimodal shift by the chimpanzee must be performed on the first try. That is, the chimpanzee must never make sounds, only utilizing visual signals. Performing the multimodal shift only after a few attempts at communication through the usual primary channel would invalidate the purpose of the experiment. This would result in a series of trial-and-error attempts, which would have nothing to do with the mind-reading hypothesis but rather with a complementary behaviour-reading hypothesis (see figure 10). For the experiment to be considered valid, the chimpanzee's multimodal shift must be the first attempt at communication with the experimenter.



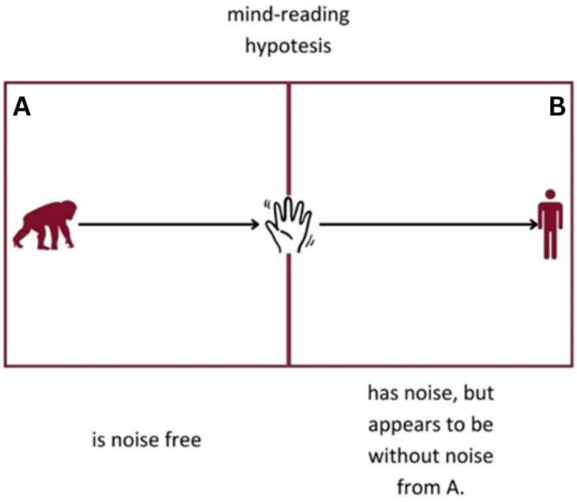


Figure 10 shows the experimental setup during the test phase. The arrow indicates the action of sending a signal to the experimenter. The hand indicates the use of a visual signal. The wavy lines indicate the use of a vocalization. According to the behaviour-reading hypothesis, the chimpanzee will not perform the multimodal shift, while according to the mind-reading hypothesis, it will.

This experiment offers an innovative approach to demonstrate the ability of perception attribution in chimpanzees, differing from traditional studies on ToM and animal behaviour. The test is not based on a single behavioural event or isolated response of the chimpanzee, as is the case in all other existing protocols, but rather on a sequence of communicative attempts that take place over an extended time span. This experiment enables the assessment of the chimpanzee's communicative behaviour in response to the perception attributed to the experimenter. It focuses on a series of communicative interactions that occur over time, providing insight into the persistence and consistency of chimpanzee communicative behaviour. The experiment avoids relying on single behavioural events or isolated responses. The experiment's core is observing how the chimpanzee chooses to use multimodal shift as its primary communication mechanism from the start of the interaction and whether it continues to use this strategy throughout the experiment. This aspect is important because consistently adapting communication to the experimenter's abilities without reverting to previously unsuccessful methods indicates a thorough understanding of their perceptions. The chimpanzee's behaviour suggests that it does not simply react to isolated stimuli or form basic cause-and-effect associations through trial-and-error. Instead, it appears to act based on a more advanced comprehension of the experimenter's perceptions and potential sensory experiences. Additionally, the experiment demonstrates the chimpanzee's capacity to integrate and apply this comprehension in a novel and previously unencountered situation. The chimpanzee's ability to recognize the unique perceptions of their experimenter is confirmed by their consistent and strategic actions during communication. This is demonstrated by their ability to communicate effectively even in noisy environments or when the caregiver is out of sight, without relying on signals that cannot be perceived by the other. In summary, the experiment is noteworthy for its ability to evaluate not only the chimpanzee's immediate comprehension of others' perspectives but also its capacity to maintain a suitable and consistent communicative approach over time. This persistence and consistency in communicative behaviour, in the absence of ineffective attempts, provides a strong indication of elements of ToM, demonstrating a sophisticated understanding of others' sensory experiences that goes beyond simply reading behaviour.

However, if persistence and consistency in communicative behaviour is the element that ensures that we can discriminate between the mindreading hypothesis and the complementary behaviour-reading hypothesis then the whole experimental procedure can be simplified by removing the element of illusory context. By removing the illusory context, the experimental procedure would work as explained above, but without the need to divide the experimental area into two separate areas. Also, the learning phase would not have to be performed. The simplified experimental procedure would work as follows. The first phase is the familiarization phase. During the familiarization phase, the chimpanzee (or another animal) is encouraged to use its preferred communication channel to interact with an experimenter and obtain food, establishing a link between communicative behaviour and reward. In this phase, the chimpanzee is asked to interact freely for a set period of time (one or more minutes) with the researcher. At the end of the set time, the researcher gives the chimpanzee food. The chimpanzee's different communicative attempts are recorded and it is determined which is the chimpanzee's preferred way of communicating (e.g., with vocalizations). This is followed by the disturbance test phase. During the disturbance test phase a disturbance element is introduced that aims to block communication through the habitual communication channel. This forces the chimpanzee to seek alternative ways to achieve the communicative goal. Again, the chimpanzee tries to get the experimenter's attention for the same amount of time established during the familiarization phase. At the end of the established time, the researcher offers him food. A control phase follows. In the control phase, two different tests can be performed. The first test is to repeat the familiarization phase and check again that in the absence of a disturbing element the chimpanzee attempts to attract the experimenter's attention by vocalizing. The second test is to repeat the familiarization phase with food present but without the experimenter present. During the second test of the control phase, the chimpanzee should either make no communicative signals, as no experimenter is present who is able to give it the food, or it should just vocalize, increasing the volume of vocalizations to try to attract the experimenter's attention wherever the experimenter is. In this simplified version of the experimental paradigm if the chimpanzee switches from the auditory communicative channel to an alternative communicative channel, without ever using vocalizations, throughout the disturbance test phase then the chimpanzee's behaviour can be interpreted according to the mind-reading hypothesis. If, on the other hand, during the set time the chimpanzee makes even one vocalization, then the chimpanzee's behaviour can be interpreted according to a behaviour-reading hypothesis. As mentioned just above, the persistence and consistency in communicative behaviour, in the absence of ineffective attempts, provides a strong indication of elements of ToM, demonstrating a sophisticated understanding of others' sensory experiences that goes beyond simply reading behaviour.

In general, an animal's ability to use an effective communication channel in any environmental context, taking into account the various disturbing elements, can only be explained as a mind-reading hypothesis and not with its own complementary behaviour-reading hypothesis. The only condition to be met is that the animal performs the correct multimodal shift, never exhibiting trial-and-error behaviour, even when it faces for the first time a new experimental situation of which it could have no previous experience. In other words, the animal must not be able to adapt its communication channel based on a previously learned cause-and-effect relationship between a given situation and a given behaviour. Such an experimental paradigm is capable of both solving the logical problem and simplifying the paradigm proposed by Lurz.

Having defined an experimental paradigm that does not run into the logic problem in testing the perceptual ToM of an animal, it is now possible to analyze how to construct an experimental paradigm that does not run into the logic problem in testing the ToM on false beliefs in non-linguistic animals.

4.3 How to solve the logic problem for false belief

In solving the logic problem regarding investigations of perceptual ToM in non-human animals, the strategy used to devise a valid experimental protocol was to start from an earlier protocol proposed by Lurz. In the case of investigations of ToM on false belief, a similar strategy is also intended to be used to solve the logic problem. However, it is first necessary to provide some background on Lurz's ARM theory and explain in detail how he hypothesizes to solve the logic problem for false belief.

To explain how ARM theory accounts for an animal's ability to attribute beliefs to others, it is important to distinguish between perceiving the appearance of an object and having beliefs about the object's actual state. According to the ARM theory, attributing beliefs to others involves understanding that the other agent has certain beliefs about the world, which may or may not be in line with objective reality. It goes beyond simply recognizing how objects appear to another agent. This process requires an animal to recognize that another agent may have representations of the world that guide its behaviour, even in the absence of direct perceptual stimuli. For instance, if an agent acts on a belief about an object that is currently occluded or not directly perceivable, the observing animal must infer that the agent's actions are guided by an internal representation or belief about the object's existence and properties, rather than by immediate sensory perception. This inference is based on the animal's ability to distinguish between appearance and reality and to understand that other agents may also make this distinction and act accordingly. For instance, if a chimpanzee observes another chimpanzee reacting to a concealed food source, it may deduce that the other chimpanzee believes that food is present there, even in the absence of direct sensory evidence. The capacity to attribute beliefs to others, as per ARM theory, is thus founded on the animal's comprehension of the difference between how things appear and how they are believed to be. This comprehension enables the animal to anticipate and interpret actions that are not directly linked to the present perceptual field, but rather rely on internal representations of the world that may stem from past experiences, acquired knowledge, or inferred circumstances.

4.3.1 Revisability belief-attribution protocol

An experiment based on the concept of revisability was conducted, taking inspiration from a study by Sato et al. (1997). The study demonstrated that chimpanzees can recognize the continuity of an object even when it is partially hidden, indicating a sensitivity to visual illusions. The Lurz test is structured as follows: in the habituation phase, chimpanzees are shown two types of scenes recorded on video. In the first video, a chimpanzee is observed moving towards a whole banana, indicating attraction. In the second video, the chimpanzee reacts to pieces of banana and then moves away, showing disinterest or repulsion towards unappealing fragments of banana. During the test phase, a more complex situation is presented where the subject is confronted with two additional recorded videos. The video depicts a chimpanzee facing an object that appears to be a whole banana, but is partially obscured, making it unclear whether it is a whole banana or two pieces. The scenario can be interpreted in two ways: either the chimpanzee sees two pieces of banana arranged behind the occluding object, or it sees a whole banana. It is important to note the ambiguity of the situation. In this case, the chimpanzee in the video is aware that what seems to be a whole, palatable banana is actually a broken, nonpalatable one. In the second video, the two pieces of banana are arranged behind the occluder before the chimpanzee enters the scene. Therefore, the chimpanzee in the video is not aware that what appears to be a whole, palatable banana is actually a broken, nonpalatable one. The experiment concludes by showing two different possible reactions in both scenarios. The chimpanzee portrayed in the video is attracted to the banana in one scenario, indicating that it believes it is a whole banana and therefore palatable. In the other scenario, the chimpanzee is not attracted to the banana, indicating that it believes it is a broken banana and therefore not palatable (see Figure 11).

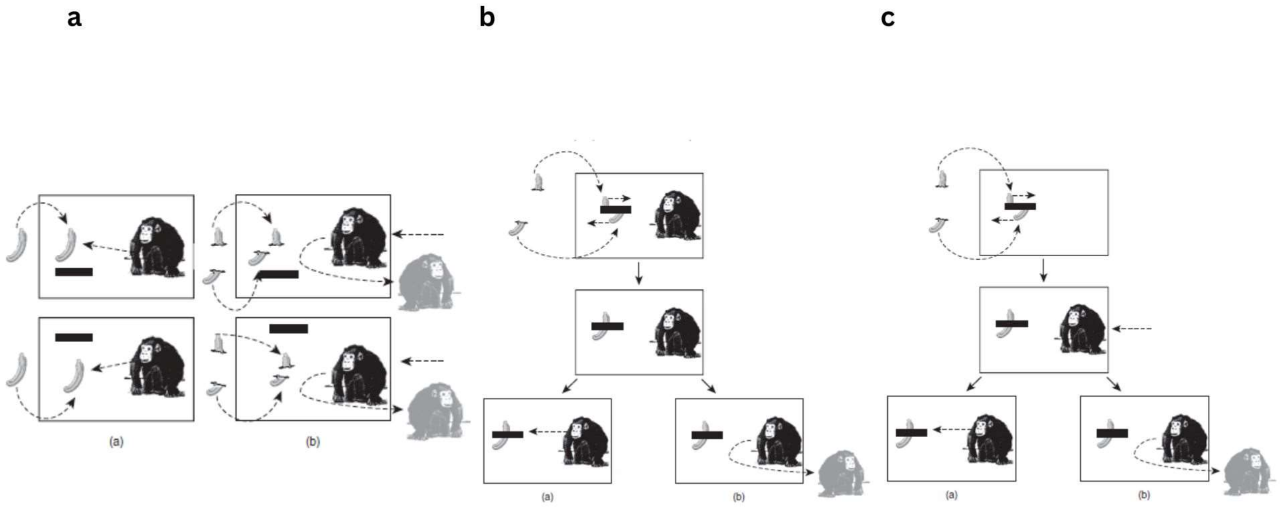


Figure 11 shows (a) a familiarization video used in two sets (a) and (b). In set (a), the video chimpanzee appears onscreen before the whole banana, and retrieves the banana when it appears. In set (b), the banana pieces appear first, and then the video chimpanzee arrives, looks at the banana pieces, and departs. (b) Testing phase 1, where the video chimpanzee is present while banana pieces are set, and (c) testing phase 2, where the banana pieces are set before the chimpanzee arrives, are also shown. This figure is reprinted from Lurz (2011).

Conclusions can be drawn from this experiment based on the reactions of the chimpanzees watching the videos. If the chimpanzees expect the chimpanzee portrayed in the video to be attracted to the banana in both scenarios, then this suggests that the chimpanzee watching the videos is basing its expectations solely on the visual appearance of the object, without considering that the chimpanzee portrayed in the video may have seen the broken banana arranged behind the occluder and therefore may know that it is not a whole banana. In this case, the chimpanzee can only attribute perceptual states and not beliefs while watching the videos. If the chimpanzee observing the videos shows surprise in certain cases, such as when it expects the chimpanzee in the video to not be attracted to the banana when it is behind the occluder, but has reason to believe that the banana is broken, this indicates that it can attribute not only perceptions but also beliefs based on previously collected information that is no longer directly perceivable.

In this experiment, as with the experiments on perception attribution, a complementary behaviour reading hypothesis is not possible. The data that the chimpanzee observing the videos relies on to create an expectation about the chimpanzee imaged in the videos differs depending on whether the chimpanzee observing the videos uses a behaviour-reading ability or a mind-reading ability to predict the behaviour of the chimpanzee imaged in the videos. In the first case, the observing chimpanzee should expect the same behaviour from the chimpanzee in each video. In the mind-reading hypothesis, however, the expectation about the behaviour of the chimpanzee in the videos differs depending on the case.

4.3.2 Abstract belief-attribution protocol

The second type of test is based on the concept of abstractness[68]. In this experiment, we analyze how the tested animals interpret the behaviour of a computer-generated image (CGI) of one of their conspecifics during two phases: the familiarization phase and the test phase. During the familiarization phase, the animal watches a series of videos in which the CGI figure stands in front of two trees and turns toward the animal that is watching the video. During the

---

[68] For this type of experiment, animals, primarily chimpanzees, undergo a preliminary Appearance-Reality test to evaluate their ability to differentiate between appearance and reality using visual or auditory amodal completion stimuli. Only those who pass this test are chosen to participate in the main experiment.

experiment, two stimuli, a red rectangle, and a blue rectangle, move silently on the screen and position themselves behind the trees, becoming occluded from view while the CGI figure is turned. The figure then turns to face the now hidden stimuli and moves along one of two paths: to the right if the stimuli are of congruent (same) shapes or to the left if they are incongruent (different). In the test phase, the animal is shown similar videos but with new and deceptive stimuli. In one scenario, two stimuli that appear identical but are actually different (e.g., an orange oval with a missing section and a complete blue oval) are placed behind trees to create the illusion of two congruent objects behind the occlusion. In the other scenario, two stimuli that are actually the same but appear different due to occlusion (e.g., two PacMan shapes that create the illusion of a complete circle and a PacMan behind the trees) are presented in the same way. After the CGI figure observes the stimuli, the test animal is presented with two possible conclusions for each scenario: a 'surprise,' where the CGI figure chooses an unanticipated path based on the perceived congruence or incongruence of the stimuli, and an 'expectation,' where it follows the expected path from the habituation phase (see figure 12). The observation time of the tested animal is measured for each conclusion. The objective of the test is to determine whether the animal being tested can comprehend that the CGI figure will select a path based on its belief about the congruence of the stimuli, despite the visual deception. If the animal being tested anticipates that the CGI figure will follow the expected path in the presence of stimuli that appear to be congruent but are not, it indicates that the animal can ascribe a belief. This means that the animal understands that the CGI figure acts based on what it believes it sees, not just what it sees directly. In contrast, if the animal does not show a significant difference in observation time between the 'surprise' and 'expectation' conclusions, it may indicate that it is limited to attributing perceptual states or its own complementary behaviour-reading hypothesis based solely on what is visually evident, without attributing more complex beliefs.

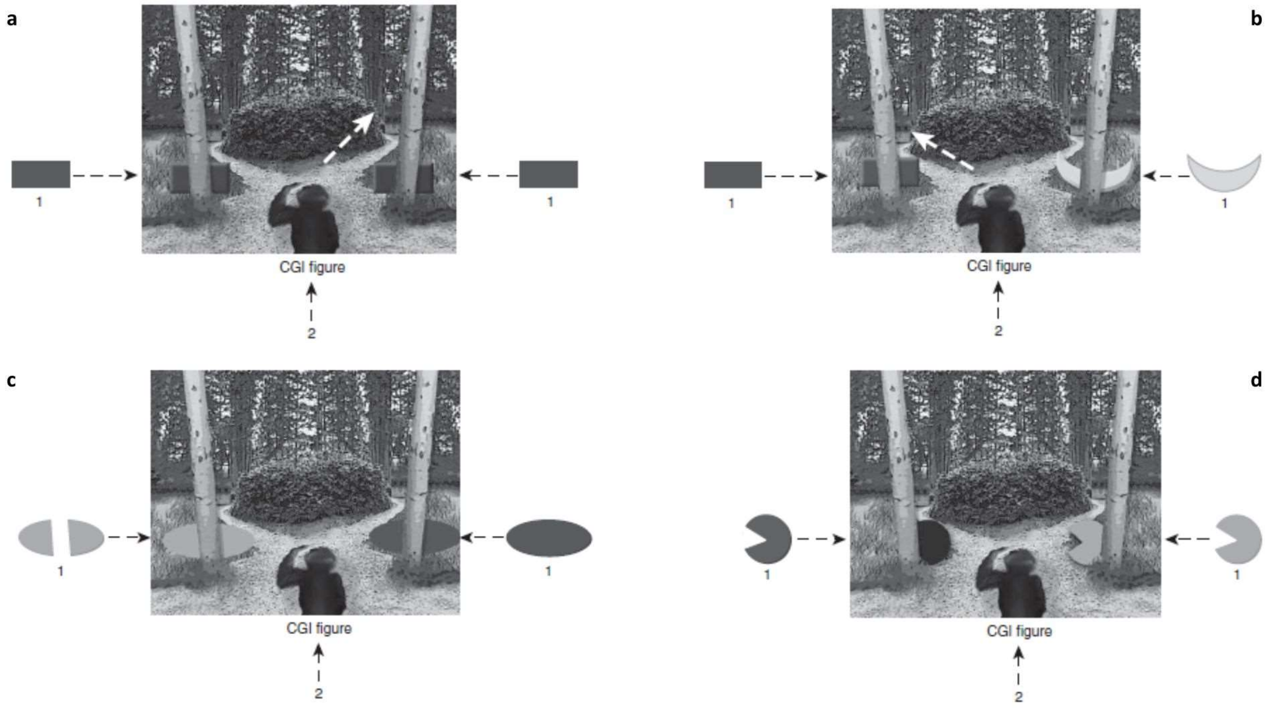


Figure 12 displays the two stages of habituation with (a) congruent and (b) noncongruent stimuli. The following are the two test phases: (c) 'look-same-but-are-different' and (d) 'look-different-but-are-same.' This figure is a reprint from Lurz 2011.

In this case, a complementary behaviour-reading hypothesis would suggest that animals could predict the behaviour of the CGI figure simply by observing and interpreting visual clues without ascribing true beliefs. However, this experiment is not easily interpreted based on reading behaviour due to the relationships between objects that are not directly observable, such as congruence or incongruence between them. Interpreting these complex relationships based solely on visual clues is insufficient. The visual appearance of objects can be deceptive, as objects may appear the same but be different or appear different but be the same. Therefore, direct observation is not enough to understand these relationships. The visual appearance of objects can be deceptive, as objects may appear the same but be different or appear different but be the same. An animal that relies solely on behaviour reading would make predictions based solely on the immediate appearance of objects, without considering additional information that might contradict these appearances. Therefore, if the animal acted according to behaviour reading, its predictions would be opposite to those of an animal acting by mind reading.

One distinguishing feature of these experiments is that the test subject, such as a chimpanzee, does not actively interact with or manipulate stimuli within the experimental setup. Instead, the subject's participation is primarily limited to observing scenarios presented via animated videos or simulated situations. In these scenarios, an agent, who may be a CGI character or a human being, interacts with various stimuli or objects. The subject's attention is measured by observing their reaction to videos, specifically their observation time towards expected versus surprising events. This reflects their ability to form expectations based on the observed behaviour of the agent in the videos. In contrast, exploiting multimodal shift allows the construction of an experimental protocol to test the ToM of false belief that is capable of solving the logic problem and at the same time involves the animal as an active part of the experiment.

4.3.3 Multimodal-shift belief-attribution protocol

The hypothetical experiment will proceed as follows: The experiment area is divided into two zones: an area labeled A (where the animal is located) and an area labeled B (where the keeper is located). Two buttons are located in area A. One of the buttons is visible to the researcher and the other hidden behind a wall that obstructs visibility. Communication area A can be disrupted by playing a disturbing sound. A chimpanzee is in area A, while a researcher is in area B. Familiarization Phase: In this phase of the experiment, a chimpanzee learn that it can obtain food from an experimenter if it presses the correct button to the type of signal the experimenter emits. At the beginning of each session both the chimpanzee and researcher explore the configuration of the experimental area. The experimental area can either exhibit area A with or without a disturbing sound. Area B, on the other hand, is always silent. Both the chimpanzee and the experimenters cannot hear any disturbing sound while it is in area B. The disturbance sound can be heard in area A, if present, and cannot be heard in area B. The researcher uses gestures when there is a noisy environment, and auditory signals when there is a silent environment. The chimpanzee learns to stand behind the correct button (behind the wall for sound, visible for gesture) before the researcher issues the signal. Making the correct choice results in the reception of food. The chimpanzee is considered to learn how to obtain food when it presses buttons correctly by anticipating the experimenter's behaviour. The chimpanzee exhibits a perceptual multimodal shift in response to environmental stimuli. When a disturbing sound is present, the chimpanzee positions itself behind the visible button, indicating its readiness to receive visual signal from the experimenter. Conversely, when the environment is silent, the chimpanzee assumes a position behind the button that is not visible to the experimenter, indicating its readiness to receive auditory signal. The experimental phase follows. During the experimental phase both the chimpanzee and researcher are made aware of a specific configuration of the experimental area. In the new configuration, area A is strongly sound disturbed. Both the chimpanzee and the experimenters while in area A can hear the disturbance sound. Area B, on the other hand, is silent. Both the chimpanzee and the experimenters cannot hear any disturbing sound while it is in area B. The disturbance sound audible in area A cannot be heard in area B. After experiencing the specific configuration of this trial, the chimpanzee goes to area A, while the researcher goes to area B. Once they are in their designated areas, the disturbing sound in area A is unexpectedly turned off without the researcher's knowledge. The chimpanzee must anticipate that even if he no longer perceives any noise, the researcher will still use a gesture instead of a sound. In fact, the researcher is unaware that the disturbing sound in area A has ceased and therefore has a false belief that area A is still noisy. If the chimpanzee understands that the researcher is unaware of the cessation of noise and expects a gesture, the chimpanzee will stand behind the visible button to see the gesture and press the correct button to obtain food. If the chimpanzee does not understand the researcher's mistaken belief and expects a sound, it may choose the wrong button and therefore not receive the food (figure 13). Control condition: In the control condition, the experiment is performed as just described in the experimental phase, but the researcher does not experience the specific configuration of the experimental area. Only the chimpanzee is made aware of the specific configuration whereby area A strongly sound disturbed, and area B silent. After experiencing the specific configuration of this trial, the chimpanzee goes to area A, while the researcher goes to area B. There are two sessions in this phase. In one session, once they are in their designated areas, the disturbing sound in area A is unexpectedly turned off. In the other session, the disturbing sound in area A is not turned off. The chimpanzee in this case should never stand behind the visible button. Whether the disturbing sound in area A ceases or remains. If the chimpanzee understands that the researcher is unaware of the disturbing sound in area A, the chimpanzee will expects a sound and it will stand behind the hidden button to try to hear the researcher' sound and press the correct button to obtain food, regardless of the presence or absence of sound in area A. If, on the other hand, the chimpanzee chooses a different button based on the presence or absence of noise, it means that it does not take into account the researcher's mental state to try to predict its behaviour.

The reason why such a designed experiment does not present the logic problem is that the chimpanzee's behavioural response is different if it has ToM or if it merely reads the researcher's behaviour. That is, there are no complementary behaviour-reading hypothesis to the mindreadind hypothesis. If the chimpanzee in fact has ToM, it will stand behind the visible button during the experimental phase; if it merely reads the behaviour then it will stand behind the hidden button.

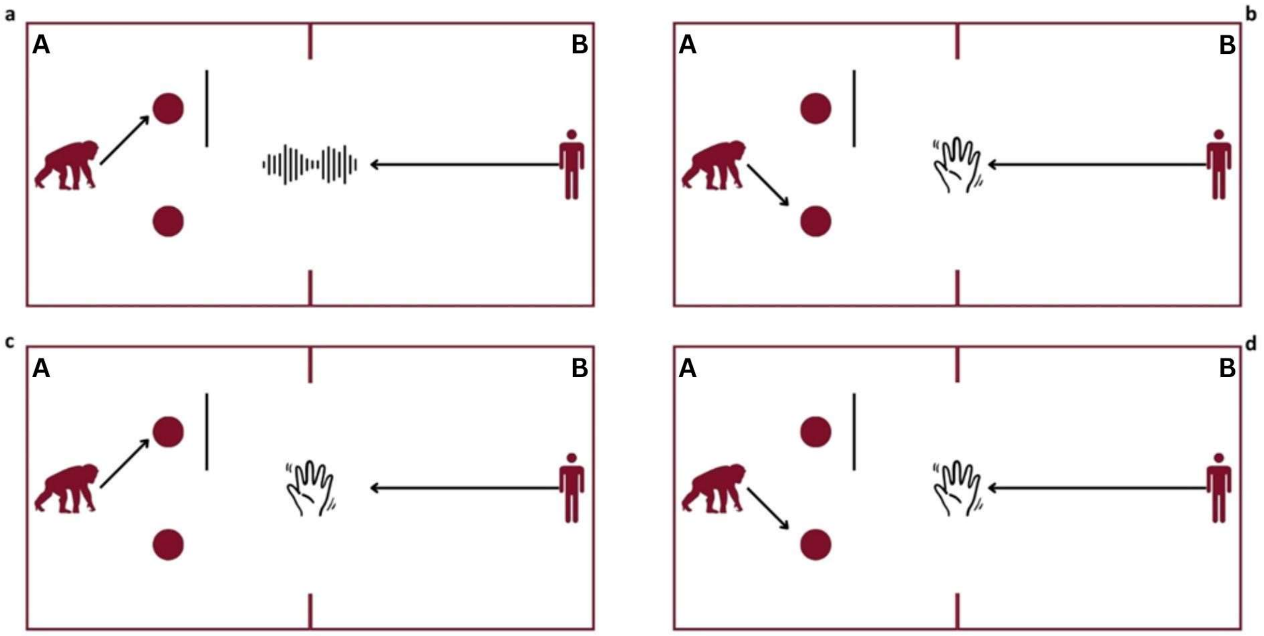


Figure 13 displays both the familiarization (a) and the experimental conditions (b). The arrow indicates the action of sending a signal to the chimpanzee. The hand indicates the use of a visual signal. The wavy lines indicate the use of a vocalization. The chimpanzee learns to position itself on the correct button based on the type of signal performed by the researcher. Chimpanzee position itself behind the hidden button (a) if the researcher produces a sound while position itself behind the visible button (b) if the researcher produces a visible signal. The behaviour-reading hypothesis and the mind-reading hypothesis are presented at the bottom. In the condition with disturbing sound, the researcher will perform a visual signal. If the chimpanzee can read the researcher's mind, it will position itself behind the visible button (d). If the chimpanzee only reads the researcher's behaviour, it will position itself behind the hidden button(c).

In a context where chimpanzees rely solely on behaviour-reading, each experiment situation would require learning a specific cause-and-effect correlation. Learning a specific cause-and-effect correlation for each experiment situation necessitates memorization of multiple discrete associations and may not be effective in novel or modified situations where such associations do not apply directly. In contrast, if the chimpanzee uses mind-reading, it applies a general principle that allows it to anticipate the researcher's behaviour in various circumstances without depending on specific pre-learned associations. This general principle is based on the understanding that the researcher will act based on their belief of the characteristics of the environment, regardless of the conditions actually perceived. The chimpanzee can predict that the researcher will use a gesture even in the absence of noise because it understands that the researcher mistakenly believes that the environment is still noisy. This ability to use mind-reading to apply a general principle to different situations demonstrates cognitive flexibility that can reduce the cognitive load of having to learn and remember multiple specific associations. It also allows the chimpanzee to adapt quickly to new situations.

This test is an improvement over Lurz's proposed protocols for measuring an animal's predictive ability. Unlike the experiments mentioned above, where the animals are primarily passive observers, this test requires the chimpanzee to act in the first person by performing active actions, such as choosing to stand behind a visible or hidden button before the researcher issues the signal. This study measures animals' predictive ability and understanding of a situation through active engagement, rather than relying on measuring their state of surprise when their expectations are violated. It is important to note that, unlike the multimodal shift test used to identify perceptual attribution ability, this study aims to investigate belief attribution ability and cannot eliminate the illusory factor from the experimental paradigm.

4.4 A general protocol to test ToM and pilot trial with two Asian elephants

This section briefly explains how the multimodal shift protocol for testing ToM in non-linguistic animals can be applied in practice. Therefore, the experiment to test perceptual ToM (section 4.2.4) and the experiment to test ToM of false belief (section 4.3.3) are combined into one protocol. Finally, a practical application of the protocol is shown with an experiment conducted on two Asian elephants. The multimodal shift protocol has been designed to be universally applicable to different animal species. Prior to testing, the animal must be able to perform multimodal shift in accordance with the characteristics specified in Chapter 3[69]. It takes into account the possibility that the animal under investigation may not exhibit a multimodal shift or that the characteristics of the shift may vary significantly from species to species. The protocol's structure comprises three main steps. The initial step is crucial in determining whether the animal under study exhibits multimodal shift behaviour and, if so, what its distinguishing characteristics are. The ability to perform a multimodal shift is considered a fundamental prerequisite for proceeding with the next steps of the protocol. The protocol proceeds by implementing a series of collaborative tasks, as detailed in Section 4.2.4, to assess the animal's ability to attribute perceptions to other individuals. This phase utilizes the observations collected during Phase 1 to test the animal's comprehension of others' sensory experiences. Finally, the protocol expands further, as discussed in Section 4.3.3, to test the animal's ability to recognize and attribute false beliefs to others. If the hypothesis that the illusory factor can be eliminated to test the ability to attribute perceptions is correct, the frist phase can fulfill both the function of identifying possible multimodal shifts and demonstrating perceptual ToM in non-linguistic animals.

The initial stage of the protocol aims to identify the presence and characteristics of multimodal shift in the animal under study. This process starts with a comprehensive review of the existing literature on communication studies related to the species in question. It is crucial to comprehend the animal's natural communication methods, both with conspecifics and, in ex situ conservation contexts, with keepers. This comprehension enables us to customize the experiment to the animal's natural communication methods, increasing the probability of observing significant behaviours. After obtaining a thorough understanding of the animal's preferred communication channels, we create a familiarization task. During this phase, the animal is motivated to use its preferred communication channel to interact with an experimenter to receive food. This task reinforces the relationship between the animal's communicative action and the reward received. It also provides a baseline for evaluating changes in communicative behaviour. In the next test phase, a disruptive element will be introduced to prevent communication through the animal's preferred perceptual channel. For

---

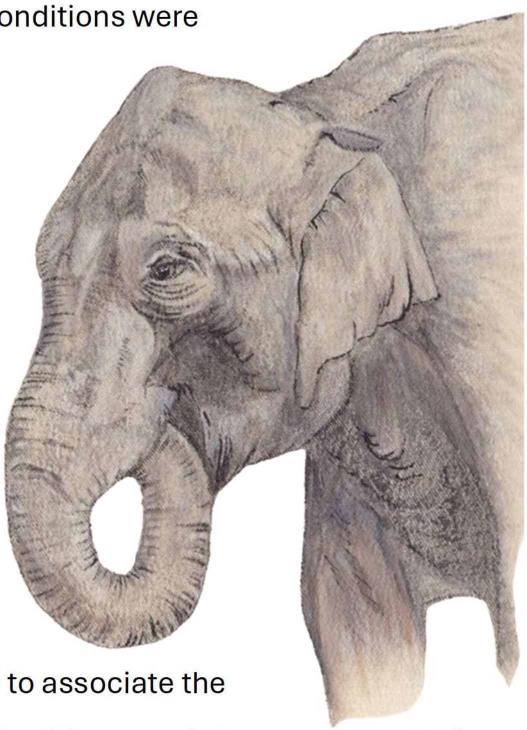
[69] A multimodal ToM-indicative shift must possess four characteristics. Firstly, it must not be an invariant response to a stimulus. Secondly, it must be based on redundant and free signals. Thirdly, it must be performed only in the presence of a perceiver. Fourthly, it must involve both the productive and perceptual aspects of communication.

instance, if the test species communicates mainly through auditory signals, a constant background noise could be introduced to hinder the animal's reliance on such signals. Likewise, for species that rely primarily on visual communication, a visual barrier could be created between the animal and the experimenter or the lighting of the environment could be altered. The key aspect of this step is to observe how the animal adapts to the disturbance. An animal that can make a multimodal shift recognizes the ineffectiveness of its usual communication channel and seeks alternatives to achieve its goal. By analyzing such behavioural adaptations, we can confirm the animal's ability to make a multimodal shift and identify which new perceptual channels are recruited to overcome the communicative obstacle. This approach provides a more profound comprehension of the flexibility and complexity of animal communication, establishing the groundwork for the subsequent stages of the protocol that concentrate on attributing perceptions and beliefs.

The second phase of the protocol focuses on exploring the animal's ability to attribute perceptions to others. This builds on the knowledge gained during the first phase regarding multimodal shift. This section of the protocol introduces an illusory element to test the animal's ability to understand others' perceptions. The animal is presented with a situation where it must recognize and react to a discrepancy between its own perceptual experience and that of another agent, usually an experimenter. This phase of the protocol corresponds to section 4.2.4. If the illusory factor is eliminated the first phase can fulfill the function of the second phase of the protocol.

The third phase of the protocol involves analyzing the animal's ability to attribute false beliefs. This builds on the previous phases, particularly the ability to make a multimodal shift and understand others' perceptions in illusory contexts. In this phase, the experiment utilizes a setup that creates a more complex illusion than the previous one. The aim is to test whether the animal can comprehend that another individual may hold a belief that does not correspond to objective reality. This phase of the protocol corresponds to the one described in section 4.3.3.

This protocol was applied in its version without an illusory element to test perceptual ToM in two Asian elephants. With the help of Professor Robert Lurz and doctoral student Vesta Eleuteri, and the support of the Rome Zoo, we have successfully implemented the first part of the protocol. Our focus was on observing multimodal shifts in two Asian elephants hosted by the zoo, namely 49-year-old Lakshmi and 53-year-old Sofia. Prior to commencing the test, the veterinary records of both subjects were reviewed to determine if any medical conditions affected their perceptual ability. No relevant clinical conditions were identified.



The initial phase of the test involved a three-day familiarization period, during which the elephants were gradually acclimated to the experimental procedure. Twice daily, an experimenter, identified as E2, stood in front of the elephants' enclosure with a box of food while elephants were present (Figure 14 condition a and condition b). After one minutes' wait, E2 offered the food to the elephant, as shown in Figure 14 (b and d). This routine was repeated twice a day during the three-day familiarization period to associate the presence of the experimenter and the food box with a positive experience and reward. The familiarization phase was recorded by an E1 experimenter for audio and video purposes. To minimize any possible elephant's expectations based on previous interactions, the experimental sessions began 48 hours after the conclusion of the familiarization phase.

We collected video and acoustic recordings of the whole reunion using an Iphone SE (2nd generation) and an omnidirectional Neumann microphone KM183 modified to record frequencies below 20Hz (flat recording: 5 Hz) and connected to a Mix Pre-6 sound device recorder at 48 kHz sampling rate. We transferred videos and audio recordings to a MacBook Pro and synchronized the separate video and audio files using DaVinci Resolve version 17.

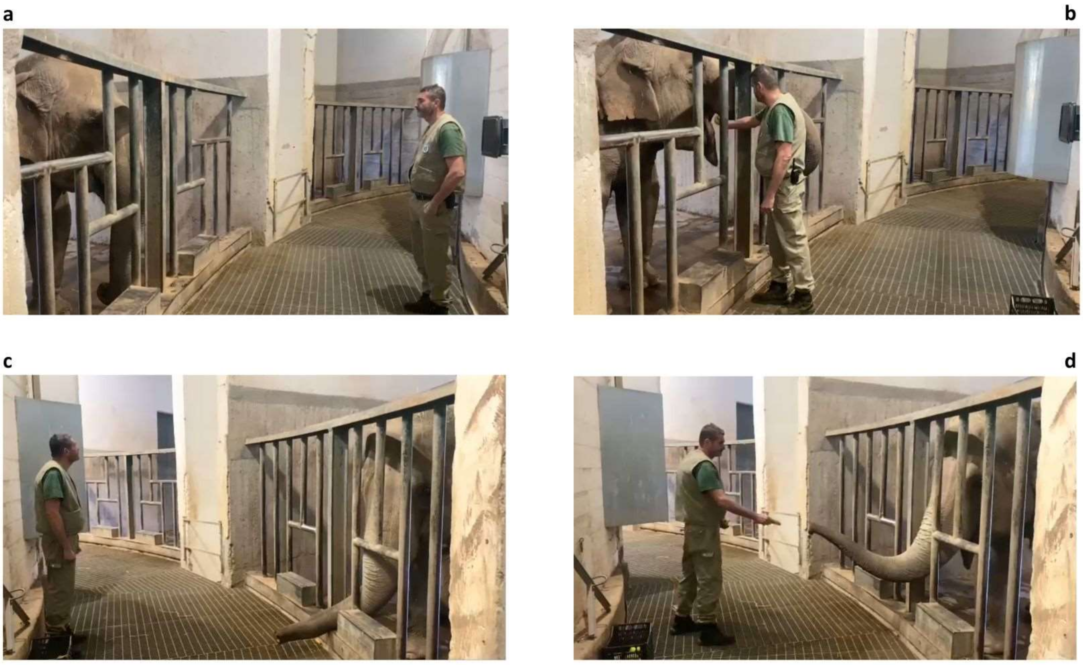


Figure 14 shows the beginning and end of the familiarization phase with Lakshmi (above) and Sofia (below). In a and c the experimenter stands in front of Lakshmi and Sofia, respectively, with the food box at his feet. In b and d the experimenter gives the food to Lakshmi and Sofia respectively.

After the familiarization trial the experiment procedure consisted of each elephant participating in a control trial an experimental trial and a third trial, each lasting one minute. The order of the sessions was balanced among the subjects. The interval between sessions ranged from 24 to 36 hours. Throughout all sessions, the elephants were separated from the human experimenters by a barrier. Both audio and video recordings are made of the sessions. In the control trial (control), experimenter E1 sets up the recording equipment and starts recording audio and video. After starting the recording, E1 calls in experimenter E2, who enters the area with a food box and places it near their feet. During this trial, E2 positions themselves to look directly at the elephant. E1 monitors the experiment's time and signals the start and end of the 1-minute session. At the end of the designated time, E2 interacts with the elephant by offering it food before leaving the experimental area. Recordings conclude at the end of the trial. In the experimental trial (condition 1), the procedure is identical to the control trial, except that when E2 enters and places the food, he turns his back to the elephant before the session begins. E1 initiates recording and signals the start and end of the 1-minute session. During the session, E2 did not respond to the elephant's gestures and remained with his back turned. At the end of the minute, E2 turned around, offered food to the elephant, and then left the area. The recordings ended at the end of the session. An interval of at least one day was left between the experimental and control sessions to minimize confusion in the elephants about the required task. The third condition (condition 2) was structured like this: E2 was positioned at a 90-degree angle to the elephant, adding complexity to the task. The expected results, according to the multimodal mind theory were the absence in visual cues in favor of auditory cues in the experimental trials compared to the control. It is possible that auditory cues would remain constant or increase, such as tapping one's foot on the floor. The multimodal mind theory also predicted that the elephant may try to capture E2's attention using alternative methods, such as spitting with its trunk.

The two female elephants in the three conditions produced different results, but they were in line with what the multimodal mind theory predicted. The exhibited signals were classified into four categories: (i) visual only, the elephant used the trunk to point to himself, the box with the food, or the experimenter (ii) sound + visual + contact, the elephant spit toward the experimenter, reaching out to him with the spit. To spit it would quickly raise its trunk and emit a loud sound (iii) visual + contact, the elephant spit toward the experimenter, reaching out to him with the spit. The elephant quickly raised its trunk but made no sound. (iv) sound only, the elephant would bang its trunk against the cage or blow noisily. In addition to the communicative signals, two different behaviours were observed: sniffing and an attempt to reach the food box. However, it should be noted that the researcher could not perceive the visual element of the signal emitted while their back was turned. A signal was considered stand-alone if it was more than one second away from the previous signal. Signals with pauses of less than one second or those occurring simultaneously were considered a unique signal. In the control condition, Lakshmi signaled once visually and three times with a multimodal signal that included visual, sound, and contact signals. The sequence of signals emitted by Lakshmi was as follows: visual; visual + sound + contact three times, then left. In condition 1, where the researcher's back was turned, she signaled only twice before leaving, using the visual + sound + contact signal in both cases. During condition 2, Lakshmi signaled four times before leaving. The signals included visual signals; then visual +sound + contact signals; and finally she walked away and produced two sound signals by tapping her trunk on the cage. In contrast, Sofia in the control condition signaled three times. She then made a visual signal, followed by a visual + contact signal, and finally an unclear signal that may have been visual or visual + contact. Before signaling, Sofia sniffed the food box three times. In the first condition, Sofia did not signal, but instead sniffed the box four times. To ensure consistency, we repeated the condition at the end of the experiment, and during the second repetition, Sofia signaled only once. First, she tried to reach the food box with her trunk, then she signaled visually + acoustically + contact. In the second condition, it signaled only twice: First she sniffed the food, then she emitted one visual and one visual +sound + contact signal and finally left. In the other conditions, the subject remained without leaving (see Figure 15).

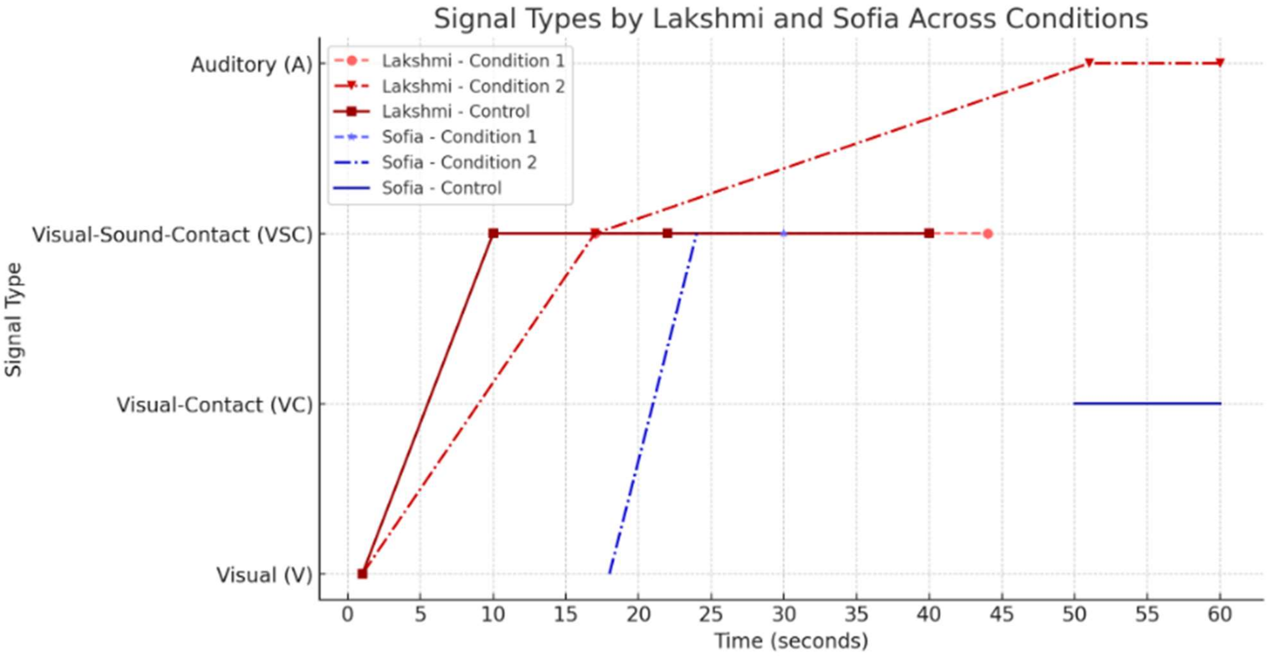


Figure 15 shows the elapsed time in each trial on the x-axis and the different types of signals on the y-axis. The resulting graph represents the sequence of signals emitted, i.e., the multimodal shift. The different points indicate when the signal was emitted, the line indicates the type of multimodal shift performed. Lakshmi signals more frequently than Sofia, but there are overlaps between their patterns. In the control condition, both Sofia and Lakshmi start with visual signals and then switch to a combination of visual + sound + contact signals. In condition 1, both Sofia and Lakshmi use visual + sound + contact signals exclusively. In condition 2, Lakshmi displays a wider range of signaling compared to Sofia, who only signals visually and with a visual + sound + contact signal.

The initial findings of the experiment suggest that elephants are capable of exhibiting multimodal shift, switching from a unimodal channel to multiple channels such as visual and contact or visual, sound, and contact. Additionally, from the researcher's perspective, the shift appears to be complete. When the researcher's back is turned, the elephant cannot observe the visual component of the multimodal signal. Therefore, the shift is from a visual signal to either a sound and contact or contact-only signal. It is debatable whether elephants have ToM, at least on a perceptual level. The elephants were subjected to a new condition with no previous experience to build on. The elephants' ability to perform a multimodal shift on the first attempt and maintain communicative consistency throughout the experiment suggests that the mechanism behind the shift is mind-reading rather than behaviour-reading. This indicates intentional communication.

However, it is important to note that these results are preliminary and derived from a limited sample of only two subjects involved in the experiment. This limitation significantly reduces the generalizability of the results and requires caution when interpreting the conclusions. To establish the presence of a ToM in elephants and the intentional nature of their communication with greater certainty, it is essential to replicate this experiment on a larger and more diverse statistical sample of subjects. Further research on a variety of individuals and in different contexts is necessary to fully understand the cognitive and communicative abilities of elephants and confirm the promising indications that have emerged from this initial study. Additionally, a more cautious approach should be taken, and the use of Lurz's protocol using illusory settings should not be ruled out in principle. If multiple tests on various individuals and animal species demonstrate that those who pass the initial step of my protocol can also pass Lurz's protocol, then we could replace the latter protocol (which is exclusively for testing perceptual order ToM) with the former. The former is simpler and more aligned with typical animal behaviour.

In the following chapter, the two primary arguments against attributing ToM to non-linguistic animals will be briefly discussed before concluding this research.

5 Examining the two main arguments against non-linguistic ToM

The previous chapter addressed the challenge of interpreting whether animals possess a ToM or whether they simply rely on behavioural information to predict the actions of others. This challenge arises from the difficulty of experimentally distinguishing between attributing mental states and reading behavioural evidence, since both interpretations are based on the same observable data. Povinelli and Vonk (2008) propose the need to develop tests that can clearly differentiate between the two hypotheses. Cecilia Heyes (1998) suggests the experience-projection protocol as a solution. However, there is still a debate about the effectiveness of existing tests in solving the logic problem. Robert Lurz criticizes current protocols for their inability to overcome the problem of mindreading and proposes an alternative approach, Appearance-Reality Mindreading (ARM) Theory. ARM Theory focuses on the ability to attribute perceptual appearance states, which may have offered significant evolutionary advantages. This ability enables animals to effectively navigate illusory environments and manipulate the behaviour of others. Building on the ARM theory, Lurz introduces an advanced version of the EP protocol to address the logic problem in ToM attribution in animals. This attempt aims to make the complementary behaviour-reading hypothesis less convincing. The modified protocol uses illusory environments to test whether an animal can understand that another's perceptions or belief may differ from its own, an ability that would involve the use of ToM. This procedure consists of three steps: creating an illusion, providing an observational learning context that excludes behaviours associated with the object's objective reality, and conducting a behavioural anticipation test to assess whether the animal anticipates specific actions by others based on the illusory perception. Despite the ingenuity of this approach, Halina raises objections based on the Hempel theorist's dilemma. She argues that any attribution of mental states could be traced to observable behavioural patterns, making ToM attribution unnecessary. Halina proposes that only a series of consistent tests in different situations could convince us of the existence of ToM, following Mill's method of agreement. Lurz acknowledges the importance of multiple testing but emphasizes the need to design experiments that isolate mental attribution abilities. A proposal was made to adapt Lurz's protocol to incorporate the multimodal shift. This offers a simplified approach that is less dependent on illusory contexts. This new paradigm involves a learning phase in which the animal associates communication with rewards, a test phase with disturbances that prevent habitual communication, and observation of how the animal adapts to these disturbances. The approach aims to test ToM in animals through their ability to modify communication strategies in response to perceptual changes, providing a potentially more direct and applicable method for assessing animals' understanding of others' perceptions. Finally, a pilot trial was conducted with two Asian elephants that suggests their ability to attribute perceptions. This chapter briefly addresses the metaphysical question that remains undetermined. The entire research has been based on the assumption that it is possible to have Theory of Mind (ToM) in the absence of language. However, as previously stated in both the introduction and Chapter 1, some argue that animals cannot attribute perceptions or beliefs to others without language. The primary arguments against ToM without language are those of Davidson (1975,1982,1997) and Bermudez (2003, 2009). Section 5.1 discusses Davidson's differentialism and the hypothesis that language is necessary for concepts and associated cognitive abilities. Section 5.2 analyzes Bermudez's view that animals may have limited rationality but cannot engage in logical reasoning or meta-representational attitudes, including ToM. In this chapter I will briefly attempt to contrast multimodal mind theory with the theories of Davidson and Bermùdez.

5.1 Davidson's argument against non-linguistic ToM

In the ongoing discussion about whether non-language animals possess ToM, the thought of philosopher Donald Davidson (1975,1982,1997) occupies a prominent position. While Davidson acknowledges a shared evolutionary history between humans and animals and is not concerned with being labeled anthropocentric or violating naturalistic principles, he strongly asserts that this similarity does not necessarily imply equivalence or direct comparisons in cognitive abilities. Davidson (1982) argues that language is necessary for thinking, but Glock (2018) notes that this only applies to intentional states and not to other cognitive processes like sensations and emotions. Although Davidson does not directly address this point, his reflections on moral conduct towards animals, especially those without language, suggest that such beings are equally deserving of kindness. This implies the presence of non-linguistic sensations and emotions (Davidson, 1982).

Davidson does not explicitly mention ToM as much as he does intentional states. which are closely related to ToM in the fields of cognitive psychology and philosophy of mind. Intentional states refer to mental states that are directed towards something outside of oneself, such as beliefs, desires, hopes, fears, and intentions. The content of intentional states can be expressed as 'X wants that Y,' 'X believes that Y,' etc., where X is the agent and Y is the content. ToM is the mechanism by which we attribute and recognize intentional states in others and ourselves. To use ToM, in other words, we must be able to conceive of others (and ourselves) as beings who have desires, beliefs, intentions, and thoughts about the world. ToM allows us to interpret or predict their behaviour based on the intentional states we attribute to them. For example, if we see a person walking toward a refrigerator, we might use our ToM to infer that the person intends to get something to eat or drink. This inference relies on the ability to attribute a specific intentional state (such as hunger or thirst) and an action directed towards satisfying that state to the person in question. In summary, intentional states are the 'contents' or 'objects' of our thoughts and feelings about the world, and ToM is the cognitive process that allows us to recognize, attribute, and reflect on such states, both in ourselves and in others. The relationship between intentional states and ToM is fundamental: intentional states are a necessary condition for the development and practice of ToM. Intentional states, such as beliefs, desires, intentions, and other forms of thinking and feeling that relate to situations, events, or actions, are necessary for the development and exercise of ToM. These intentional states are directed towards something and have an object or content on which they focus. To possess ToM, the ability to attribute mental states to oneself and others in order to interpret and predict behaviour, it is essential to be able to conceive of such intentional states.

After clarifying this point, we can summarize Davidson's argument into three interconnected points that support his thesis: intentional states require the presence of language. The first test, known as the 'intentionality test,' asserts that we cannot attribute specific intentional states to nonhuman animals without a shared language that allows us to understand how they conceive of a given situation. Instead, we can only hypothesize among various possible interpretations without being able to determine the animal's conception precisely. Davidson illustrates this point with a thought experiment, inspired by Malcolm, in which a dog is chasing a cat toward a tree. When the dog isn’t looking, the cat scurries away but not up the tree. The dog, not seeing this, runs to the tree, looks up into the tree, and begins barking. Davidson argues that attributing beliefs to animals is not necessary to explain their behaviour. Therefore, we cannot assume that the dog actually believes the cat is in the tree, but rather that this attribution helps us understand the dog's behaviour. Similarly, we cannot say that a missile desires to destroy an airplane, but rather that it is programmed to follow and intercept it. Furthermore, attributing intentional states to a dog would imply the attribution of concepts such as 'cat' and 'tree.' This presupposes a capacity for perception and description that a dog, lacking language, cannot express. This argument is based on what Glock refers to as the 'lingualist master-argument,' which holds that thought requires concepts that depend on language. According to this view, thought is intrinsically linked to language. Therefore, nonhuman animals, lacking language, would not be capable of thinking.

Davidson does not explicitly deny the presence of conceptual thoughts or intentional states in nonhuman animals, contrary to the interpretation of some of his critics. Instead, he questions our ability to accurately articulate and describe these mental states in animals using specific terms known as 'de dicto' descriptions. These descriptions refer to how a subject conceives of a situation, including the specific beliefs and thoughts they have about it. Davidson raises doubts about our ability to precisely express what animals think or believe. However, some criticisms have emerged regarding the difficulty of distinguishing between the various possible 'de dicto' beliefs of nonhuman animals. These criticisms suggest that with advances in behavioural and empirical knowledge about animals, we may be able to more clearly identify their mental states in the future. In addition, some alternative theoretical approaches, such as the one proposed by Bermúdez (section 2.5), use a modified version of 'success semantics.' This approach considers beliefs as causal factors of animal behaviours aimed at satisfying their desires, thus offering a new way of interpreting animal mental states. However, Armstrong (1973) points out that, in Malcolm's example of the attribution of intentional states, we should deal with 'de re' attributions and not 'de dicto' attributions. "De re" attributions describe the objective situation to which the subject reacts, without considering the specific way in which the subject mentally represents that situation (de dicto). For example, when we say that the dog believes or is convinced that the cat is in the tree, we are referring to the objective situation (de re). The focus shifts from the conceptual nature of the dog's thinking to the concrete object of its perceptions and reactions. This change in focus emphasizes the dog's observable behaviour rather than its internal mental processes.

Davidson's second argument, known as the 'argument from holism,' deals with the attribution of 'de re' intentional states. The argument suggests that attributing intentional states to a non-linguistic animal based solely on perceptions, without subjective interpretation, would be impractical. Davidson argues that for an intentional state to be possessed in relation to a given situation, it must be part of a larger network of related intentional states. For instance, in order for the dog in Malcolm's example to believe that the cat climbed the tree, it must have a set of other beliefs related to the cat, the tree, the action of climbing the tree, and the surroundings in general. Only by comprehending this entire network of interconnected beliefs can we truly understand what the dog perceives. However, it is impossible to attribute 'de re' intentional states to nonhuman animals, such as the dog in Malcolm's example, without the ability to explore the existence and structure of their belief network from the outside. This is further complicated by the absence of a language that allows animals to express their intentional states. Fodor and Lepore (1994) criticize the argument, stating that it leads to the absurd conclusion that intentional states cannot be attributed to even humans who share the same language, thus leading to the absurdity that we cannot attribute intentional states to either nonhuman animals or our fellow humans.

Davidson's third argument, which builds on the conclusions of the first two, is the most complex and central. The first two arguments focus on the impossibility of attributing intentional states to non-linguistic animals, but do not deny the possibility that they possess them. The third argument posits that nonhuman animals, due to their lack of language, cannot participate in cognitive processes involving intentional states, whether de dicto or de re. Davidson's argument is based on two premises: firstly, possessing intentional states requires having the concept of intentional states, and secondly, language is necessary to acquire such concepts. Davidson focuses on the phenomenon of surprise, which occurs when a belief held to be true is discovered to be false. This realization involves a comparison between intentional states and the situation itself, requiring an understanding of intentional state and truth and falsity concepts. According to the holism argument, intentional states must be embedded in a network of connected intentional states. This network includes the idea that these states can be true or false with respect to a given objective reality. Davidson argues that the ability to conceptualize the objectivity of a situation derives from a capacity for triangulation. According to Davidson, triangulation is a process that involves three key points: the subject, the object or situation, and another subject. The subject is the person who is thinking or perceiving, the object or situation is what the subject is thinking about or perceiving, and the other subject is another person who can perceive the same situation or object and with whom the first subject can communicate. This process is fundamental to establishing the common objectivity of the world and to developing the concepts of truth and falsity. To illustrate, consider a scenario where two individuals, Alice and Bob, are observing a tree. Alice believes the tree is a beech tree, while Bob believes it is an oak tree. In this case, the tree is the object or situation that they are both perceiving. Their disagreement prompts them to compare and discuss their perceptions and beliefs about the tree, thereby involving language. During the discussion, Alice and Bob may point out specific features of the tree, such as the shape of the leaves or the structure of the bark, to support their respective beliefs about its identity. Through this exchange, they use language to express their intentional states and attempt to reach an agreement about the objective nature of the tree. Triangulation occurs when Alice and Bob interact with both the object (the tree) and each other, using language to share and compare their beliefs and perceptions. This process helps establish a shared understanding of reality and develop concepts of truth and falsehood regarding the tree. Davidson argues that language-mediated interaction enables individuals to develop complex concepts, such as intentional states, and attribute objective meanings to their experiences. This ability is not present in nonhuman animals, which lack language.

In short, Davidson argues that language is essential for forming and communicating concepts, such as those involved in the attribution of beliefs and intentions, and without it, we cannot attribute specific intentional states to nonhuman animals. Moreover, attributing "de re" intentional states to nonhuman animals is impractical without a common language, since such states require a network of related beliefs that only language can express and communicate. This is because understanding and ascribing such states requires the concept of intentional states themselves and the ability to triangulate, both of which depend on language.

Several researchers, such as Tye (1997), Carruthers (2008), and Lurz (2009, 2011), have examined Davidson's theses and raised objections. They argue that possessing the concept of an intentional state is not essential to experiencing an intentional state, and that this process can occur even without conscious awareness. On the other hand, they challenge the idea that Davidson's first two arguments demonstrate the impossibility of attributing intentional states to beings without language. Instead, they propose that these arguments only highlight the difficulty of justifying such attributions. Carruthers, in particular, challenges Davidson's interpretation of the phenomenon of surprise. According to the author, surprise arises from a mechanism that detects conflicts between different concepts associated with intentional states. These conflicts activate physical reactions, and the perception of these reactions constitutes the experience of being surprised. This experience does not require a complex conscious process or language. Furthermore, similar to how we attribute intentional states to nonhuman animals based on their behaviour, animals should be able to do the same with other beings. This suggests that nonhuman animals may have access to the capacity for triangulation, which Davidson considers to be dependent on language. This perspective is supported by empirical data suggesting that animals are capable of attributing intentional states to others. It offers a more inclusive view of animal cognitive abilities.

Davidson's arguments are compared to those of Stich (1979) and Dummett (2010), who also conclude that it is impossible to attribute beliefs to non-linguistic entities. However, they make significant changes to the original theoretical framework and enrich it. Unlike Davidson, who emphasizes the need for beliefs to be true, Stich places emphasis on the sharing of the belief network rather than its truthfulness. Stich argues that knowledge of the entire belief network is not necessary to attribute a specific belief, as in the case of Malcolm's dog believing the cat in the tree. From this perspective, the main difference between Davidson and Stich is that Davidson argues against animals participating in intentional states, while Stich suggests that we lack a solid epistemological basis to determine whether animals possess such capacities. Newen and Starzak (2022) argue that intentional states do not always require an infinite network of related intentional states. Instead, they can exist within networks of intentional states of varying extents. This suggests that nonhuman animals could also participate in intentional states with less complex networks than humans.

Newen and Starzak introduce two important counterarguments. These concern the ability to interpret the behaviour of nonhuman animals by attributing intentional states to them without necessarily fully understanding their content and the possibility of approaching such understanding in an explanatory and justified way. The first argument is based on the principle that behaviour can be explained as the result of the interaction between informational states, such as beliefs, which provide information about a specific situation, and motivational states, such as desires, which define how one would like that situation to be. This distinction highlights that it is not sufficient to know a single belief to explain a behaviour, but it is necessary to consider the set of beliefs and desires related to that situation. Beliefs should be understood as 'decoupled representations' rather than automatic responses to specific stimuli, which Millikan (1996) refers to as 'pushmi-pullyu representations[70]' that have both an informational and motivational function. The key is to establish behavioural criteria for distinguishing whether certain nonhuman animals engage in pushmi-pullyu representations or intentional states. Newen and Starzak suggest analyzing broader behavioural patterns rather than isolated behaviours. This approach can test whether an animal's representations of a state of affairs are tied to a single behavioural response or can be integrated with other states to generate different and novel behaviours. The authors suggest that behavioural flexibility may indicate the presence of intentional states in nonhuman animals, even without a full understanding of their contents.

Newen and Starzak utilize the concept of behavioural flexibility as a sufficient indicator for identifying the presence of intentional states in non-linguistic animals, without the need to access the precise content of those states. They do not consider it a necessary condition for having intentional states. This perspective differs from the notion that beliefs and intentional states should have a specific and well-defined content, and a form of representation that

---

[70] Ruth Millikan's representations, known as 'pushmi-pullyu', are named after a fictional creature in Hugh Lofting's children's book 'Dr. Dolittle'. The creature has two heads, one of a goat and one of a unicorn, located at opposite ends of its body, allowing it to move in opposite directions simultaneously. Millikan employs the metaphor of pushmi-pullyu representations to describe a type of representation that combines two functions: descriptive and imperative. A pushmi-pullyu representation provides information about the environment or a state of affairs, which is descriptive in nature as it aims to reflect reality by providing data about the world. For instance, an animal may have a descriptive representation of the location of food. The representation also includes an 'imperative' component that guides or motivates action. This aspect instructs the organism on how to respond to the received information. For instance, the representation of food position can motivate the animal to move towards that position to eat. Pushmi-pullyu representations are unique because they combine these two aspects into a single cognitive entity. According to Millikan, this type of representation is essential for nonhuman animals' cognition, enabling them to act effectively in their environment without the need for abstract reasoning or explicit deliberation. These representations allow animals to react immediately and adaptively to their perceptions, integrating their understanding of the environment with motivations for action.

demonstrates their properties of compositionality, systematicity, and productivity, which are characteristic of human language. Newen and Starzak propose a more inclusive approach, citing Beck's (2012) criticism of the tendency to rationalize linguistic representations a posteriori as the only valid models and Allen and Bekoff's (2007) suggestion that the linguistic representation of intentional states may be a human-specific manifestation of a broader cognitive faculty shared with nonhuman animals. The authors argue against overemphasizing differences or similarities between humans and other animals, as this could lead to erroneous conclusions or improper generalizations. They caution that behavioural flexibility should not be the sole criterion for attributing intentional states to nonhuman animals, but rather a way of assessing their ability to form decoupled and combinable representations. This capacity varies among animal species and reflects varying degrees of cognitive complexity. The theory of intentional states proposed by Newen and Starzak aligns with Davidson and Stich's holism. However, it distinguishes itself by providing a more detailed explanation of how represented information affects behaviour. This allows for a more nuanced and detailed understanding of intentional states in humans and other animals.

Newen and Starzak's second argument aims to justify the rough attribution of content to intentional states. This approach contrasts with Stich's and builds on the ideas of Schwitzgebel (2015). Schwitzgebel points out that our ability to explain human behaviour through intentional states is inherently limited by the need to balance referentially opaque (de dicto) structures with referentially transparent (de re) structures. This requires a constant balancing act of approximating the content of intentional states without capturing their true essence. Attributing intentional states to nonhuman animals presents a challenge, particularly in translating these states from a non-linguistic to a linguistic format due to the obstacle of indeterminacy of translation, as proposed by Quine. However, Newen and Starzak argue that a similar approach can be legitimately applied to nonhuman animals, given that even when attributing intentional states to our fellow humans, we often resort to a process of approximation by a posteriori rationalization. Newen and Starzak propose three criteria for justifying the attribution of certain content to the intentional states of nonhuman animals. Firstly, the individual must exhibit sufficiently complex behaviour and demonstrate specific behavioural dispositions. Secondly, activities within this behavioural pattern must be based on informational states that can be combined with each other. Finally, the individual must exhibit faculties to react to new information, classify it, and update its informational states according to these new acquisitions. These faculties imply at least the presence of elementary memory and learning capacity.

According to Newen and Starzak's criteria, a non-linguistic animal can develop structured representations of objects and their characteristics within a given context. These representations allow the animal to associate specific objects with certain properties and organize these associations systematically. Newen and Starzak describe cognitive abilities as 'epistemic dispositions' that vary in complexity depending on an animal's sensitivity to environmental information, its ability to categorize contextual items and their characteristics, and its effectiveness in integrating this information into a belief construct. These aspects influence behavioural patterns that can be interpreted as based on intentional states, allowing for rough inferences about their contents. More advanced 'epistemic dispositions' indicate a greater ability of the organism to update its intentional states in response to new information, following the principles of correspondence (the adjustment of beliefs to observed reality) and coherence (the harmonious integration of new beliefs with pre-existing ones). Non-linguistic animals with advanced dispositions demonstrate cognitive flexibility and adaptability, making them legitimate subjects for the attribution of intentional states and the approximate understanding of their contents.

Following Newen and Starzak's arguments, it is worth noting that their thesis on the possibility of non-linguistic animals having intentional states fully supports multimodal mind theory. The three criteria for attributing intentional states to non-linguistic animals are fully met by the thesis that cross-modal binding acts as a catalyst for ToM and linguistic and communicative abilities, as well as by the phenomenon of multimodal shift. Cross-modal binding enables the integrated and unified perception of the world by combining the various sensory properties of objects, such as color, sound, shape, and texture, into a single cognitive experience. This process occurs without the need for conscious naming or categorization through language. For instance, an apple can be recognized through its shape, red color, and crisp sound when bitten into, even before the word 'apple' is known or used. It is argued that as cross-modal binding develops and is refined over time, the ability to form increasingly complex and detailed relationships between objects and their properties also increases. This process of sensory integration provides a rich cognitive foundation on which to build concepts, even before these relationships are made explicit and structured through language. According to

Beck (2012) and Allen and Bekoff (2007), the translation of non-linguistic mental content into linguistic mental content occurs only after the fact and may be a human-specific ability.

Therefore, the notion that behavioural flexibility is an indicator of intentional states only enhances the theoretical foundation for which multimodal shift behaviour, characterized by significant flexibility, could be an indicator of ToM. The multimodal shift behaviour demonstrates communication flexibility and the ability to assess and respond to a dynamic context, which requires mental representation of the context and goals. Integrating multimodal shift as an indicator within Newen and Starzak's framework could provide a more detailed perspective on intentional states and ToM in animals. This approach aligns behavioural variability with cognitive complexity and provides a concrete criterion for assessing the presence of intentional states and ToM without relying exclusively on language.

Cross-modal binding not only enables the formation of generic concepts, but also the formation of the concepts of truth, falsehood, and belief. And this contrasts with Davidson's claim. In Davidson (1982), the author presents a series of arguments aimed at establishing the improbability of non-linguistic animals possessing beliefs. These arguments, namely the 'opacity' argument, the 'holism' argument, and the 'main' argument, constitute the crux of Davidson's discourse. Central to his exposition is the articulation of his principal argument, delineated by two pivotal premises:

P1. The possession of beliefs necessitates the conceptual grasp of belief itself, along with subsequent comprehension of truth and falsity.

P2. The acquisition of concepts pertaining to truth, falsity, and belief is contingent upon engagement in linguistic communication.

The logical inference derived from P1 and P2 is that non-linguistic animals lack the requisite cognitive framework for understanding truth, falsity, or belief, thereby precluding the attribution of beliefs to them. This line of reasoning bears significance within the domain of ToM in non-linguistic animals animals, as Davidson espouses the 'primacy of belief' thesis, asserting that all mental states, including desires and intentions, are intricately intertwined with beliefs. Consequently, the absence of belief conception impedes the formation of other mental state attributions, culminating in the absence of a fully-fledged ToM in non-linguistic animals. While Davidson does not furnish explicit justification for his 'primacy of belief' thesis, he offers rationale in support of P1 and P2 of his primary argument. Notably, his defense of P1 revolves around the notion that the possession of beliefs entails susceptibility to surprise, which, in turn, necessitates holding beliefs concerning the potential falsity of one's own beliefs. Conversely, his rationale for P2 hinges upon the contention that linguistic communication serves as the primary avenue for acquiring the conceptual apparatus of truth, falsity, and belief.

The multimodal mind theory incorporates the concepts of redundant and nonredundant signals (section 2.3) and multimodal syntax, pragmatics, and syntactics. As elucidated in section 3.3, when an animal, human, or nonhuman is confronted with two pieces of information that are non-redundant and contradictory to each other, it is compelled to discriminate between the two. To do so, it must assume that one of the two pieces of information is more likely to be true than the other. In order to discriminate between the two pieces of information, the animal must rely on its previous experiences and the current state of affairs, that is, the context within which it obtains the two conflicting pieces of information. The concept of truth or falsehood emerges from the contrast of conflicting information and the consequent discovery that when two pieces of information are conflicting, only one of them corresponds to a certain state of affairs, while the other does not. The greater the number of instances in which an animal deals with conflicting information, the greater its ability to abstract the concept of truth or falsity. It is therefore the concepts of truth and falsity that contribute to the formation of the concept of belief, rather than the other way around. In short, multimodal mind theory negates both premises of Davidson's main argument by substituting cross-modal binding for language in its role regarding concept acquisition.

5.2 Bermudez’s argument against non-linguistic ToM

One recent argument against the ToM in non-linguistic animals is proposed by José Luis Bermùdez (2003, 2009). Bermùdez's argument can be divided into two sections: the first provides solid reasons to support non-linguistic thinking in non-human animals, while the second exposes the limits of such thinking. According to Bermùdez, a theory of non-linguistic thought must be able to accommodate four dimensions[1]. Therefore, an argument can explore these dimensions. A proper explanation of non-linguistic thought must be able to explain its metaphysics and semantics, including their nature, mechanisms, structure, and which thoughts can be engaged by non-linguistic creatures. Additionally, it should provide a valid methodology and psychological explanation of behaviour, allowing such thoughts to be attributed as the cause of the behaviour of non-linguistic creatures and to understand how and whether such creatures represent both their environment and the thoughts of other creatures.

In the constructive part, Bermùdez distances himself from theories of non-linguistic thought that he calls minimalist or deflationary. He argues that these theories are unable to provide an adequate explanation of certain behaviours, such as planning or intelligent action (e.g., tool use), enacted by different non-linguistic animals. These behaviours cannot be explained by the conjunction of mere motivational and perceptual states. Therefore, Bermùdez proposes a modified version of the mental content theory known as success semantics[71]. This theory suggests that the content of a belief is determined by the success conditions for the performance of an action triggered by this belief. The proposed modification allows for the ascription of thoughts to non-linguistic creatures. However, it does not provide a means of distinguishing between different thoughts that can be attributed to a creature in a specific situation. To address this issue, Bermudez combines success semantics with Quine's reification concept. Reification refers to the process of a creature transitioning from experiencing a state of affairs as a set of features to experiencing the same state of affairs as a set of subjects. Bermudez argues, in contrast to Quine, that this process is not solely dependent on language. According to Bermùdez, the combination of these two concepts enables overcoming the problem of indeterminacy in attributing thoughts to creatures present with the basic version of success semantics. This argument suggests that non-human animals can have beliefs about certain states of affairs and engage in primitive reasoning forms that use those beliefs. As a result, they can attribute perceptual states or simple goal-directed desires to other creatures.

---

[71] An approach to the theory of meaning that follows an insight of F. P. Ramsey, who points out (in 'Facts and Propositions') that we might say that a chicken believes that a kind of caterpillar is poisonous if the chicken's actions 'were such as to be useful if, and only if the caterpillars were actually poisonous.' The approach cements together the likelihood of satisfying a desire with the truth of the belief on which the agent acts. It is thus a way of developing a kind of pragmatism.

In the destructive part, Bermúdez sets the limits of non-linguistic thinking by arguing that any kind of meta-representational thinking[72], which is dependent on the capacity for semantic ascent[73], cannot be engaged in by non-human animals and other non-linguistic creatures. Bermùdez does not want to deny the ability to think in toto. However, he starts from Andy Clark's definition (1998, 2005) of language as a cognition-enhancing animal-built structure. Bermùdez argues that the six original cognitive capacities catalyzed by language can also leverage non-linguistic processes[74]. The latter are understood as processes that do not rely on linguistic symbol systems or any kind of symbols at all. The unique aspect of human cognition, only attainable through language, is the ability to engage in what Clark (1996) refers to as second-order cognitive dynamics. These capacities involve self-evaluation, self-criticism, and finely honed remedial responses.

Bermùdez (2003) defines the characteristics that make a language based on Pierce's Theory of Sign. He notes that symbols, unlike icons, are arbitrary and conventional. For instance, he argues that the bee communication system falls into the latter category. The distinction between linguistic and non-linguistic symbol systems lies in the ability of a language to form complex symbols from simple symbols, to which a truth or falsehood value can be attributed. This requires the presence of complex symbols composed of at least one naming symbol and one predicate symbol. The compositionality principle involves a complex symbolic system that can be divided into two subcategories: sequentially complex symbol systems and hierarchically complex symbol systems. The main distinction between the two is that the logical and semantic relationship among the components of a hierarchically complex symbol cannot be inferred from its components sequential order alone. Hierarchically complex symbol is created through mechanisms other than the mere combination of a predicate symbol and one or more naming symbols.

The second characteristic of the boundary line is that a suitable language must provide valid means for thoughts to become the objects of other thoughts, which must be on a personal level, rather than a sub-personal one. Bermùdez argues that only public language sentences

---

[72] All thinking that involves thinking about thoughts.

[73] The capacity to think about words.

[74] The only difference between linguistic and non-linguistic processes in terms of the cognitive capacities catalyzed would be a quantitative rather than qualitative difference.

possess the necessary characteristics for second-order cognitive dynamics, as conscious observation and control of one's own thoughts are not possible under the language of thought hypothesis, which is a hypothesis about how cognitive processes operate at a sub-personal level. Moreover, it is important to use public language sentences because alternative vehicles, such as those related to mental model theory and the idea of mental maps, cannot be inferentially connected to each other. It is worth noting that Bermùdez acknowledges the existence of thoughts formulated as mental models or mental maps, but emphasizes that these should not be the sole object of further thought. He denies that non-linguistic vehicles can convey propositional thoughts.

Bermudez argues that, besides second-order cognitive dynamics, other types of thinking require the capacity for meta-representation. This capacity can be called explicit when it directly rests on first-order thoughts. All forms of psychological explanation of behaviour, including the forms of higher order desire, fall under this category. It can also be present implicitly in some processes that do not directly rest on first-order thoughts, such as logical reasoning capable of exploiting connectives and propositional true-functional quantifiers. In summary, non-linguistic animals are incapable of engaging in logical reasoning, monitoring the formation of their own beliefs, or attributing thoughts to other animals. They only have access to analogous non-linguistic thought forms that can sustain cognitive capacities that are precursors to our own, in a rather complicated way. The cognitive divide between creatures with language and those without is significant. However, it represents a division between two modes of thinking - two ways of conceptualizing the social and physical environment - rather than a dichotomy between thought and the lack of it.

To address criticisms, Bermudez (2017) proposes a revision of his main argument to clarify his position. He proposes a twofold distinction for classifying meta-representations: metacognition, the ability to monitor and control one's own mental states, and mind-reading, the ability to think about the mental states of others. This classification divides metacognition and mind-reading into perceptual and propositional attitudes, distinguishing between linguistic and non-linguistic thinking. The distinction between perceptual states and propositional attitudes lies in the fact that the former have direct implications for action, while the latter's implications are more accurately attributed to the set of beliefs rather than a single belief. Finally, there is a difference between the consequences of misjudgment at the perceptual level versus the intentional level. In the former case, being prone to a certain kind of error can be considered a positive factor, while in the latter case, the consequences are more serious. Bermudez aims to present a picture of the union resulting from this dual classification. Assuming that non-human animals are capable of meta-representational thinking, they would only have access to the perceptual form of meta-representation for both metacognition and mind-reading, not the one related to propositional attitudes, which is reserved for humans.

The cognitive effort required for the two forms of meta-representation differs. When engaged in perceptual mind-reading, a subject must first represent another subject, then the environment that the other is representing, and finally represent to itself that the other is able to perceive that particular environment. Upon further examination, Bermudez argues that this cognitive ability may not even be considered meta-representational, as it does not necessitate the subject to represent representations to themselves, but only to represent a particular state of affairs in the world. This text discusses mind reading related to propositional attitudes, which involves representing a subject as having true or incorrect representations about their environment, rather than simply perceiving it.

Once again, the importance of language is emphasized here for two reasons: the need for belief representations to be in linguistic form and the requirement for that form to be in natural language. Bermudez assumes that representing a belief requires representing the inferential relations between that belief and others. Language is necessary for this representation because it can create complex structures from simple elements according to combinatorial rules. Additionally, specific symbols like quantifiers and logical connectives are used, which pictorial/imagistic representations lack. The second perspective assumes that mind-reading related to propositional attitudes is an integral part of conscious decision-making. This is because our decisions, in both collaborative and competitive environments, are influenced by the propositional attitudes of others. Bermudez argues that in order for this to occur, the symbols conveying this type of representation must be accessible to consciousness. This invalidates the notion that a language of thought can perform the same function.

The two aforementioned reasons as why language is crucial aims to address the comments made by John Heil (2012) and Elizabeth Camp on the use of images or maps as representations, as well as the comments made by Lurz on the language of thinking and the value of a priori reasoning. Heil argues that inner utterances used for thinking are a tool within the category of mental images. In this case, audible images are the counterpart to what we hear, just as pictorial images are to what we see. He emphasizes that there cannot be a conceptual leap between propositional images and other types of mental images. In this sense, language is viewed as a tool for thinking. Bermudez's central claim can be reformulated to suggest that thinking requires linguistic imagery. The area of study related to how we think through inner dialogue is interesting to explore and still requires further research. However, not all types of imagery can represent inferential connections between different thoughts. Therefore, non-linguistic imagery is not fit for purpose. Bermudez argues that Heil does not provide good arguments to the contrary.

The distinction between linguistic and imagistic representations can be likened to the difference between digital and analog representations. Digital representations rely on symbols that have an arbitrary connection to the object they represent and are discrete signals, while analog representations have a similarity or isomorphism relationship to what they represent and are continuous signals. Bermudez argues that purely analog symbols cannot support mind-reading related to propositional attitudes. However, he acknowledges that he did not consider the hypothesis of hybrid representations between analog and digital symbols in his original argument. This hybrid representation could potentially support thinking about thinking. He addresses Elizabeth Camp's hypothesis that diagrams and maps are sentences written in a unique notation. Camp argues that these representations differ from pictorial representations in their level of abstraction and are capable of representing negative information. Cartographic representations would enable the manipulation of basic symbols to attribute arbitrary semantic properties, similar to those found in a sentence structure. Bermudez distinguishes between implicit and explicit mastery, where the latter is required for mind reading regarding propositional attitudes. Explicit mastery is understood as the ability to explain the rules governing the combination of complex symbols. Bermudez reiterates that while it is possible to think implicitly with a map, thinking about how another creature represents itself in the world requires a language capable of explicitly communicating how the map functions as a tool for representing the world.

Bermùdez denies the possibility that the language of thought can support mind-reading related to propositional attitudes due to the necessity of explicit communication. Robert Lurz contests this assumption, considering it at least dubious. He argues that thoughts and their relations to each other must be consciously accessible, not their vehicles, allowing for remaining at a subpersonal level. Bermùdez agrees with Lurz that the vehicles of thoughts cannot be cognitively accessible, which is the main point of the debate. However, he believes that this only applies to first-order thoughts, which have a state of affairs as their object, and not to second-order thoughts, which have other thoughts as their object. If the vehicles of both categories of thought were considered cognitively inaccessible, the distinction between first- and second-order thoughts would be lost. Bermùdez's argument implies that while it is possible not to know the vehicles of first-order thoughts, these vehicles must be known for second-order thoughts. To be known, they must be in a public language. Lurz adds a methodological note to this discussion regarding Bermùdez's argument. Lurz points out that Bermùdez's argument is a priori, which means that empirical experiments aimed at verifying the possibility of mind-reading related to propositional attitudes are futile. This leaves the debate on an exclusively theoretical level. However, Lurz considers this consequence to be wrong in principle and argues that certain types of investigation cannot be performed by the armchair. Empirical investigations are needed to continue the debate. Bermúdez, on the other hand, uses Quine's metaphor of science as a force field. He believes that there cannot be an experimental protocol that alone directs the questions regarding the attribution of second-order cognitive capacities to nonhuman animals. At the same time, his proposed theoretical framework should be understood as a contribution to be included in a broader multidisciplinary context. This context has the task of keeping in balance both the different theoretical proposals and the different empirical results to produce as complete an understanding of the topic as possible.

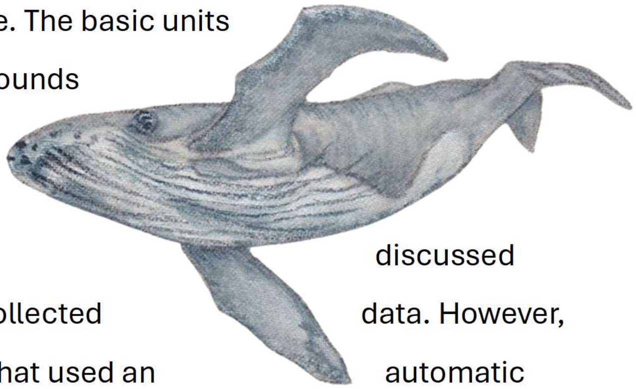
However, Bermùdez only considers cases that support his thesis, neglecting the broader multidisciplinary context, especially its empirical part. Pessi Lyrra (2005) notes in her review of Thinking Without Words that Bermùdez fails to include all relevant cases that could challenge his thesis. While the results of experiments with apes and baboons may be subject to multiple interpretations, it is recommended to apply Morgan's canon. Additionally, experiments with infants under 14 months of age suggest their ability to recognize intentions (Meltzoff et al., 1995). Call et al. (2004) and Varley (1998) successfully demonstrated the possibility of non-linguistic meta-representational thinking through the use of a non-linguistic version of the false belief test and an agrammatic aphasic subject, respectively. However, the question of what form these thoughts should take remains open. Lyrra proposes that non-linguistic conceptual thoughts connected with mental images may serve as vehicles for such forms of thought.

Lyrra argues that concepts can be conceived independently of linguistic structures, which diverges from traditional conceptions found in analytic philosophy and Bermúdez's theories. This idea finds support in cognitive sciences, where conceptual thinking is not necessarily bound to language. It is important to note that this is a subjective evaluation by both philosophers and should be clearly marked as such. Bermúdez, on the other hand, argues that possessing a concept implies the ability to think about its basis in linguistic terms. Instead, the author argues that concepts can be represented and understood through non-linguistic abilities, such as recognition. In this sense, a concept could be understood as the ability to recognize an object or entity at different times without the need for linguistic labeling. For instance, the concept of a 'chair' can be recognized without the use of language. Even without using the word 'chair,' we can recognize various objects as chairs based on common features such as legs, a seat, and a back. This ability to recognize shows that we can possess and use concepts without reflecting on them in linguistic terms. Lyrra criticizes Bermúdez's view of nonconceptual content, which is defined as content that can be attributed to a being without attributing to it the possession of the concepts involved. According to Lyrra, this distinction introduces a circularity in the acquisition of concepts and moves away from the everyday understanding of what a concept is. Additionally, Lyrra compares recognizing capacities to Bermúdez's higher-order principles. These principles help us perceive objects as whole and distinct entities. However, according to Lyrra, they should not be considered transient and limited to the duration of perception. For instance, the ability to recognize an apple as such at different times is not based on a new conceptual elaboration for each perception, but on a coherent recognition that transcends individual perceptual experiences. Lyrra notes that Bermúdez refers to non-linguistic concepts when discussing 'protonegation' in non-linguistic creatures, using the concepts of presence/absence. This seems to contradict his definition of concept possession, which would require theoretical reflection, thus undermining his own argument. In conclusion, Lyrra distinguishes the conceptual domain from the nonconceptual domain by its ability to abstract from finely detailed representational content. This allows for the identification, categorization, and recognition of objects as particular. This perspective on concepts, which can exist at both personal and sub-personal levels, facilitates metacognition even in non-linguistic creatures. It emphasizes the availability of appropriate psychological concepts rather than the possession of suitable vehicles for metacognition.

Considering both Bermùdez's argument and the critiques presented to him, it is proposed to make some considerations. Bermùdez's argument relies entirely on his personal definition of language. He argues that language can only be spoken of in the presence of complex symbols that allow the attribution of a truth value to the propositions expressed, meaning that language can only be spoken of in the presence of compositionality. According to Bermùdez, there are two variants of compositionality. The first pertains to sequentially complex systems of symbols, which is characteristic of some communicative forms of human beings. The second variant, which is exclusive to human beings, pertains to hierarchically complex symbol systems. Bermùdez argues that only hierarchically complex symbol systems are capable of supporting meta-representations and logical inferences (2003, 2017). Bermùdez also denies that animals possess language. Bermùdez does not rely solely on a unimodal definition of signs when proposing the distinction between hierarchically complex and sequentially complex symbols. However, Bermùdez suggests that compositional and recursive symbol systems may exist in a multimodal dimension. Additionally, as Lyrra has pointed out, Bermùdez fails to cite a range of empirical evidence that conflicts with his thesis. Two pieces of evidence are particularly interesting.

The initial study examines research conducted on humans in 2012 by Frank et al. The study aimed to investigate how language, which is hierarchically structured, is processed cognitively. Frank and his colleagues' research focused on analyzing language processing and highlighted how individuals can interpret and construct sentences using cognitive mechanisms that do not rely exclusively on complex hierarchical structures. The study analyzes language constructions, such as units of form and meaning, to show how sentences and words are combined in a sequential logic. This combination of constructions to form sentences may occur through parallel sequential processes, rather than hierarchical processing. This model suggests that language understanding can utilize surface clues and contextual information, reducing the importance of hierarchical structure. This fact is important for cognitive ethology because it challenges the traditional view that the complexity of human language derives exclusively from complex hierarchical structures that are distinct from the communication forms of other animals. This perspective highlights the importance of sequential structures and suggests that there may be more similarities between human and animal communication than previously assumed.

While wanting to leave out the need for a hierarchically complex symbol system for the development of meta-representations, if engaging in ToM requires such a system, are we certain that no animal species besides humans possess such a hierarchically complex symbol system? Studies have investigated the presence of hierarchically complex symbol system in primates and birds (Genter et al. 2006, De Vries et al. 2008; Ferrigno 2022). However, some of these studies may have issues with data interpretability and are often disputed. Nevertheless, the case of humpback whales (previously identified in Chapter 3 as potential ToM carriers) is emblematic. In 1971, Payne and McVay analyzed the structure of humpback whale songs based on direct field observation. Payne and McVay aimed to identify identical sound elements in the humpback whale song to determine if it uses a recursive grammar and presents hierarchically complex symbol systems. The songs, according to research, follow a hierarchical and distinct structure. The basic units of a song are unitary and interrupted emissions of sounds that persist for up to several seconds. These units can be modulated in frequency or even in amplitude. The McVay and Payne study was initially discussed based on McVay and Payne's interpretation of the collected data. However, in 2006, Suzuki and colleagues conducted a study that used an automatic classifier to group the various units of humpback whale song. The analysis confirmed that the song is an ordered sequence of individual sound elements separated by gaps of silence. It also demonstrated the existence of a strong structural constraint in song generation, occurring periodically in groups of 6 to 8 units up to 180 to 400 units thus confirming the presence of a recursive grammar in whale songs. Given that humpback whales, along with humans, possess a hierarchically complex symbol system, one may wonder why Bermùdez does not argue that Theory of Mind (ToM) is also possible for other animals besides humans?



Framing the observations made by Lurz and Lyrra within the multimodal mind theory presented in this thesis weakens the plausibility of Bermùdez's thesis. Lurz and Lyrra argue that having concepts and engaging in ToM attitudes may depend on unconscious processes. Additionally, Lyrra and Lurz both argue that concept formation is not dependent on linguistic structures. In chapter three, we discussed how cross-modal binding enables the formation of concepts even before the development of a conscious component and justifies the assumption that concept formation, as well as more complex behaviours such as ToM, can exist in their non-linguistic version, thus supporting both Lurz's and Lyrra's arguments.

Specifically, the multimodal mind theory is at odds with Bermudez's assertion that a full development of ToM, which encompasses false beliefs, is contingent upon the possession of language. Bermudez posits that language, structured in a hierarchically complex symbol system, provides the requisite vehicles for thought to manipulate representations in order to represent the representations of others. Bermudez then argues that this process must be a conscious one. The multimodal mind theory, in contrast, posits that it is cross-modal binding, an unconscious process, that allows for the structured hierarchization of multiple information as it develops. Language or different communication systems may be indicative of the level of development of cross-modal binding, exhibiting within them the same hierarchical or recursive features. However, language development is not exclusively dependent on cross-modal binding. It is therefore possible that language may be less developed than cross-modal binding, but this does not preclude cross-modal binding from structuring information according to what was called multimodal syntax and semantics in Chapter 3. A language or communication system that is not particularly developed might fail in the task of bringing complex processes that occur at the unconscious level to the conscious mind, such as the process of understanding false beliefs. Bermudez is correct in his assertion that a hierarchical structure of thought is necessary for the highest degree of ToM development. However, he is incorrect in his claim that such a structure must be found in the language or communication systems of nonhuman animals and that this structure must be conscious.

In conclusion, it is important to reiterate the points made by Lurz, as well as in the introduction and Chapter 1 of this research, regarding Bermùdez's thought. Any a priori theory concerning the nature of mind and thought are dependent on the scientific community's ability to produce empirical evidence that cannot be interpreted in more than one way. The theory of multimodal mind provided the necessary tools to solve the logic problem and brought preliminary evidence that, at least as far as attribution of perceptions is concerned, this can be found in the two elephantesses tested in the Rome Zoo. Currently, there is no firm evidence that animals other than humans can attribute beliefs. Although I disagree with Bermùdez's thesis that ToM is language-dependent, it is important to note that he does not deny the possibility of a non-linguistic animal attributing perceptions to others. However, he does deny the possibility of a non-linguistic animal attributing beliefs. Currently, the available empirical evidence does not completely contradict his assertion.

Conclusions

Before presenting the conclusions, let us summarize the research work. A historical and critical examination was conducted to explore the metaphysical and epistemological aspects of ToM research. This examination recognized the interdependence between language and communicative systems, studied through a unimodal approach, and ToM. The study of language and communicative systems has been criticized for its exclusive focus on unimodal approaches, which can lead to research bias. This bias can have implications on both metaphysical and epistemological levels. On the metaphysical level, it can lead to the attribution of unique features to language, such as recursiveness and compositionality, and the claim that these features are the reason why humans can engage in ToM attitudes, thus excluding other animals from engaging in similar cognitive capacities. The construction of experimental protocols capable of exploiting the full range of behaviours available to test animal cognitive abilities is hindered by the lack of overall understanding of animal behaviour on an epistemological level.

The study of language and other communicative systems has predominantly followed a unimodal approach since the early philosophers. This approach has influenced the study of cognitive abilities in nonhuman animals as well as disabled or neurodivergent humans. The study of multimodality gained prominence in the early 21st century and revealed two noteworthy phenomena: the multimodal shift and the correlation between an individual's ability to integrate multiple signals and their level of ToM development. Building on these two considerations, we propose a theoretical framework called multimodal mind theory, and an experimental paradigm for testing the presence of ToM in non-linguistic animals. Multimodal mind theory argues that language and ToM are not interdependent and emphasizes the role of unconscious, non-linguistic processes in supporting the hypothesis that ToM is possible in non-linguistic animals. The experimental paradigm, which investigates the presence of ToM in non-linguistic animals by exploiting the phenomenon of multimodal shift, seems to be able to overcome the logical problem of double interpretation of results and is reflected in the pilot experiment conducted on the two Asian elephant females at the Rome Zoo.

Finally, the previous chapter discussed the two main theses contrasting the idea that it is possible to engage in ToM attitudes in the absence of language: Davidson's theory and Bermùdez's theory. After comparing and criticizing both theses, it appears that multimodal mind theory finds support. Multimodal mind theory appears to provide a better explanation than Davidson's and Bermùdez's theses for some of the mechanisms underlying ToM functioning and the historical association with language.

Although the main purpose of the thesis was to integrate studies on multimodality with studies on animal cognition, in particular, to offer a viable methodology to overcome the logic problem and to lay the foundation for a new line of research regarding ToM, I believe this purpose has been achieved. However, there are still some considerations to be made.

The first consideration pertains to the question of whether non-linguistic animals can possess ToM. While multimodal mind theory supports this hypothesis, it is important to emphasize that metaphysical questions about the nature of the mind are secondary to empirical questions. Although an empirical solution has been proposed for testing ToM in non-linguistic animals, the proposed experimental protocol has not yet been fully executed, and certainly not on a sufficient number of individuals, to consider answering the question of whether non-linguistic animals have ToM or not. This remains an open question that requires investigation, particularly due to the ethical implications it carries. Both Lurz's protocol and the one proposed in this paper are valuable tools for ethologists to use in investigating the cognitive abilities of the animals they study.

The second consideration does not apply to nonhuman animals. According to multimodal mind theory, the development of cross-modal binding ability is crucial for the development of ToM ability and language. Furthermore, the theory suggests that there is no dependent relationship between language and ToM. Based on these considerations, research can be conducted to verify the validity of these arguments. This research has noted, in particular, that language, ToM, and cross-modal binding are deficient in individuals with ASD and LD, depression, anxiety, and schizophrenia. It is possible that ToM and language rely on cross-modal binding ability. Therefore, a protocol could be developed to stimulate this ability and test whether improving cross-modal binding ability leads to detectable improvements in language or ToM tasks. Each of these conditions has specific characteristics that prevent generalization of a protocol. However, this area of research shows promise and could lead to new protocols for curing or improving these conditions. Additionally, it could validate, invalidate, or provide new information to complement and refine multimodal mind theory.

The final consideration, however, pertains to the differentiation between language and communication, or between language and other forms of communication. This will be briefly addressed in light of the preceding discussion. In general, when attempting to differentiate between language and communication, two main arguments are often presented: one related to the characteristics of language and other communication systems, and the other related to the functions that language and communication serve. With respect to the first argument, language and other communication systems differ from one another in terms of discreteness, grammar, recursiveness, and displacement. Discreteness refers to the existence of individual units that can be combined. Grammar provides a system of rules that dictate how to combine individual units. Recursiveness allows for the creation of an infinite number of messages using language. Displacement enables communication about things that are not physically present or occurring in the present moment. Regarding the issue of language and its function, it can be understood either as a specific tool for human communication or as a broader cognitive faculty, as Clark notes in Bermùdez's argument. This faculty is capable of enhancing a range of non-linguistic cognitive abilities. In contrast, communication systems serve only the purpose of facilitating communication.

The argument is that the classification and definition of language, communication, and communicative systems become inapplicable when a unimodal approach is replaced by a multimodal approach. In terms of distinguishing language from other communication systems, multimodal signals can combine to form categories of redundant and non-redundant signals. Redundant signals can be enhancing, equivalent, or antagonistic, while non-redundant signals can be independent, dominant, modulating, or emergent. Is the fact that signals from different sensory channels can combine to produce a complex signal, and the way these signals combine produces the effects just mentioned, clear evidence that a form of discreteness and grammar, albeit in a much lower gradualness than human, is also present in the communication forms of non-human animals? Next, let us consider displacement. Deception mechanisms in nonhuman animals often involve influencing the behaviour of another animal, such as by falsely communicating the presence of a predator. Recursiveness has been observed in humpback whales, indicating its presence in species beyond humans. It is not intended to deny a Darwinist approach or the manifestation of these features in varying degrees. However, if the question only concerns the distinction between communication systems and language based on gradualness, then that distinction loses its meaning.

The function of communication systems and language, according to multimodal mind theory, is to catalyze the development of other cognitive capacities. This is supported by Clark's argument that language, along with other communication systems, has this function. For instance, it has been hypothesized that language and other communication systems may contribute to bringing the mechanism of cross-modal binding to a conscious level, which is claimed to occur unconsciously. According to Andy Clark, language plays a vital role in improving cognitive abilities such as memory, environmental categorization, coordination, and cooperation. These faculties are also enhanced by the communicative systems of nonhuman animals, albeit to a limited degree.

If language and other communication systems share similar features and functions, why not recognize language in non-human animals? This could be achieved by identifying species-specific configurations of language features that enhance cognitive abilities. However, this is only a consideration and requires further exploration. According to this perspective, language is a faculty that exists in varying forms and degrees across all animals. It is not solely a tool for communication, but rather communication is just one of the functions that can be accomplished through language. If applicable, language can be defined as the cognitive ability that provides the necessary tools for learning the species-specific rules for communication within a given speech community. This is subject to the rules of neuronal plasticity and can function as a scaffold for other non-linguistic cognitive abilities.

Considering these three factors, I believe that the multimodal mind theory outlined above should be further developed. This is especially important due to its potential to create new avenues of research into ToM and the relationship of this ability, along with language, to cross-modal binding, and the implications that such research might have for our understanding of the animal mind more broadly.

Bibliography

Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge. Annual Review of Psychology, 60(1), 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Ahmed, F. S., & Stephen Miller, L. (2011). Executive Function Mechanisms of Theory of Mind. Journal of Autism and Developmental Disorders, 41(5), 667–678. https://doi.org/10.1007/s10803-010-1087-7

Akçay, Ç., & Beecher, M. D. (2019). Multi-modal communication: Song sparrows increase signal redundancy in noise [Preprint]. Animal Behaviour and Cognition. https://doi.org/10.1101/750307

Allan, K. (A c. Di). (2013). The Oxford handbook of the history of linguistics (1st ed). Oxford University Press.

Allen, C., & Bekoff, M. (1999). Species of mind: The philosophy and biology of cognitive ethology. MIT.

Allen, C., & Bekoff, M. (2007). Animal Minds, Cognitive Ethology, and Ethics. The Journal of Ethics, 11(3), 299–317. https://doi.org/10.1007/s10892-007-9016-5

Andrews, E. F., Pascalau, R., Horowitz, A., Lawrence, G. M., & Johnson, P. J. (2022). Extensive Connections of the Canine Olfactory Pathway Revealed by Tractography and Dissection. The Journal of Neuroscience, 42(33), 6392–6407. https://doi.org/10.1523/JNEUROSCI.2355-21.2022

Andrews, K. (2005). Chimpanzee Theory of Mind: Looking in All the Wrong Places? Mind and Language, 20(5), 521–536. https://doi.org/10.1111/j.0268-1064.2005.00298.x

Andrews, K. (2015). The animal mind: An introduction to the philosophy of animal cognition (pp. x, 232). Routledge/Taylor & Francis Group.

Armstrong, D. M. (1973). Belief, Truth and Knowledge. Cambridge University Press.

Astington, J. W., Harris, P. L., & Olson, D. R. (A c. Di). (1988). Developing theories of mind. Cambridge University Press.

Auersperg, A. M. I., Von Bayern, A. M. P., Gajdon, G. K., Huber, L., & Kacelnik, A. (2011). Flexibility in Problem Solving and Tool Use of Kea and New Caledonian Crows in a Multi Access Box Paradigm. PLoS ONE, 6(6), e20231. https://doi.org/10.1371/journal.pone.0020231

Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. Trends in Cognitive Sciences, 6(1), 47–52. https://doi.org/10.1016/S1364-6613(00)01819-2

Bailey, A., Phillips, W., & Rutter, M. (1996). Autism: Towards an Integration of Clinical, Genetic, Neuropsychological, and Neurobiological Perspectives. Journal of Child Psychology and Psychiatry, 37(1), 89–126. https://doi.org/10.1111/j.1469-7610.1996.tb01381.x

Baillargeon, R. (1995). Physical reasoning in infancy. In The cognitive neurosciences (pp. 181–204). The MIT Press.

Balgova, E., Diveica, V., Jackson, R. L., & Binney, R. J. (2023). Overlapping Neural Correlates Underpin Theory of Mind and Semantic Cognition: Evidence from a Meta-Analysis of 344 Functional Neuroimaging Studies [Preprint]. Neuroscience. https://doi.org/10.1101/2023.08.16.553506

Baracchi, D., Cabirol, A., Devaud, J.-M., Haase, A., d'Ettorre, P., & Giurfa, M. (2020). Pheromone components affect motivation and induce persistent modulation of associative learning and memory in honey bees. Communications Biology, 3(1), 447. https://doi.org/10.1038/s42003-020-01183-x

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: Cognitive or affective? Journal of Autism and Developmental Disorders, 18(3), 379–402. https://doi.org/10.1007/BF02212194

Baron-Cohen, S. (1989). Joint-attention deficits in autism: Towards a cognitive analysis. Development and Psychopathology, 1(3), 185–189. https://doi.org/10.1017/S0954579400000377

Baron-Cohen, S. (1989a). The Autistic Child's Theory of Mind: A Case of Specific Developmental Delay. Journal of Child Psychology and Psychiatry, 30(2), 285–297. https://doi.org/10.1111/j.1469-7610.1989.tb00241.x

Baron-Cohen, S. (1989b). The theory of mind hypothesis of autism: A reply to Boucher. International Journal of Language & Communication Disorders, 24(2), 199–200. https://doi.org/10.3109/13682828909011956

Baron-Cohen, S. (1995). Mindblindness: An Essay on Autism and Theory of Mind. The MIT Press. https://doi.org/10.7551/mitpress/4635.001.0001

Baron-Cohen, S. (2000). Theory of mind and autism: A review. In International Review of Research in Mental Retardation (Vol. 23, pp. 169–184). Elsevier. https://doi.org/10.1016/S0074-7750(00)80010-5

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? Cognition, 21(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and Intentional understanding of picture stories in autistic children. British Journal of Developmental Psychology, 4(2), 113–125. https://doi.org/10.1111/j.2044-835X.1986.tb01003.x

Barrett, J. L., Richert, R. A., & Driesenga, A. (2001). God's Beliefs versus Mother's: The Development of Nonhuman Agent Concepts. Child Development, 72(1), 50–65. https://doi.org/10.1111/1467-8624.00265

Beck, J. (2013). Why we can't say what animals think. Philosophical Psychology, 26(4), 520–546. https://doi.org/10.1080/09515089.2012.670922

Bekoff, M. (1972). The Development of Social Interaction, Play, and Metacommunication in Mammals: An Ethological Perspective. The Quarterly Review of Biology, 47(4), 412–434. https://doi.org/10.1086/407400

Belwood, J. J., & Morris, G. K. (1987). Bat Predation and Its Influence on Calling Behaviour in Neotropical Katydids. Science, 238(4823), 64–67. https://doi.org/10.1126/science.238.4823.64

Bennett, J. (1978). Some remarks about concepts. Behavioural and Brain Sciences, 1(4), 557–560. https://doi.org/10.1017/S0140525X00076573

Bensoussan, S., Cornil, M., Meunier-Salaün, M.-C., & Tallet, C. (2016). Piglets Learn to Use Combined Human-Given Visual and Auditory Signals to Find a Hidden Reward in an Object Choice Task. PLOS ONE, 11(10), e0164988. https://doi.org/10.1371/journal.pone.0164988

Berardi, N., Pizzorusso, T., & Maffei, L. (2000). Critical periods during sensory development. Current Opinion in Neurobiology, 10(1), 138–145. https://doi.org/10.1016/S0959-4388(99)00047-1

Bermúdez, J. L. (2003). Thinking without Words. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159691.001.0001

Bermúdez, J. L. (2017). Can Non-linguistic Animals think about Thinking? In K. Andrews & J. Beck (A c. Di), The Routledge Handbook of Philosophy of Animal Minds (1ª ed., pp. 119–130). Routledge. https://doi.org/10.4324/9781315742250-12

Białecka, M., Wodniecka, Z., Muszyńska, K., Szpak, M., & Haman, E. (2023). Both L1 and L2 proficiency impact ToM reasoning in children aged 4 to 6. Painting a more nuanced picture of the relation between bilingualism and ToM. Bilingualism: Language and Cognition, 1–19. https://doi.org/10.1017/S1366728923000652

Bourjade, M., Dafreville, M., Scola, C., & Jover, M. (2023). Six-month-old infants' communication in a comparative perspective: Do maternal attention and interaction matter? Journal of Experimental Child Psychology, 231, 105651. https://doi.org/10.1016/j.jecp.2023.105651

Braastad, B. O., McBride, A., & Newberry, R. C. (2022a). The Cat: Behaviour and Welfare. CABI.

Braastad, B. O., McBride, A., & Newberry, R. C. (2022b). The Cat: Behaviour and Welfare. CABI. https://doi.org/10.1079/9781789242331.0000

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. Developmental Psychology, 18(6), 906–921. https://doi.org/10.1037/0012-1649.18.6.906

Brothers, L. (1990). The neural basis of primate social communication. Motivation and Emotion, 14(2), 81–91. https://doi.org/10.1007/BF00991637

Brumm, H., & Slabbekoorn, H. (2005). Acoustic Communication in Noise. In Advances in the Study of Behaviour (Vol. 35, pp. 151–209). Elsevier. https://doi.org/10.1016/S0065-3454(05)35004-2

Buckner, C. (2014). The Semantic Problem(s) with Research on Animal Mind-Reading: The Semantic Problem(s) with Research on Animal Mind-Reading. Mind & Language, 29(5), 566–589. https://doi.org/10.1111/mila.12066

Bulgarelli, D., Testa, S., & Molina, P. (2022). Theory of Mind Development in Italian Children with Specific Language Impairment and Autism Spectrum Disorder: Delay, Deficit, or Neither? Journal of Autism and Developmental Disorders, 52(12), 5356–5366. https://doi.org/10.1007/s10803-021-05389-8

Burge, T. (2007). Foundations of Mind. Oxford University Press.

Burge, T. (2010). Origins of Objectivity. Oxford University Press.

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett, M. (2003). Neural correlates of cross-modal binding. Nature Neuroscience, 6(2), 190–195. https://doi.org/10.1038/nn993

Bushmann, P. J. (1999). Concurrent Signals and Behavioural Plasticity in Blue Crab (Callinectes sapidus Rathbun) Courtship. The Biological Bulletin, 197(1), 63–71. https://doi.org/10.2307/1542997

Byrne, R. W. (1995). The thinking ape: Evolutionary origins of intelligence. Oxford University Press.

Byrne, R. W., & Whiten, A. (1988). Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans (pp. xiv, 413). Clarendon Press/Oxford University Press.

Byrne, R. W., & Whiten, A. (1990). Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans. Behaviour and Philosophy, 18(1), 73–75.

C. L. M. (1926). Chimpanzee Intelligence and its Vocal Expressions. By Robert M. Yerkes and Blanche W. Learned . (Baltimore: The Williams and Wilkins Company. 1925.). Philosophy, 1(01), 114–115. https://doi.org/10.1017/S0031819100014868

Cai, H., & Dent, M. L. (2020). Attention capture in birds performing an auditory streaming task. PLOS ONE, 15(6), e0235420. https://doi.org/10.1371/journal.pone.0235420

Call, J., & Tomasello, M. (1994). Production and comprehension of referential pointing by orangutans (Pongo pygmaeus). Journal of Comparative Psychology, 108(4), 307–317. https://doi.org/10.1037/0735-7036.108.4.307

Call, J., & Tomasello, M. (1999). A Nonverbal False Belief Task: The Performance of Children and Great Apes. Child Development, 70(2), 381–395. https://doi.org/10.1111/1467-8624.00028

Calvert, G. A. (2001). Crossmodal Processing in the Human Brain: Insights from Functional Neuroimaging Studies. Cerebral Cortex, 11(12), 1110–1123. https://doi.org/10.1093/cercor/11.12.1110

Carruthers, P. (2000). Phenomenal Consciousness: A Naturalistic Theory. Cambridge University Press.

Carruthers, P. (2008). Meta-cognition in Animals: A Skeptical Look. Mind & Language, 23(1), 58–89. https://doi.org/10.1111/j.1468-0017.2007.00329.x

Carruthers, P. (2009a). How We Know Our Own Minds: The Relationship Between Mindreading and Metacognition. Behavioural and Brain Sciences, 32(2), 121–138. https://doi.org/10.1017/s0140525x09000545

Carruthers, P. (2009b). Invertebrate concepts confront the generality constraint (and win). In R. W. Lurz (A c. Di), The Philosophy of Animal Minds (pp. 89–107). Cambridge University Press. https://doi.org/10.1017/CBO9780511819001.006

Carruthers, P. (2013). Mindreading in Infancy. Mind & Language, 28(2), 141–172. https://doi.org/10.1111/mila.12014

Catala, A., Mang, B., Wallis, L., & Huber, L. (2017). Dogs demonstrate perspective taking based on geometrical gaze following in a Guesser–Knower task. Animal Cognition, 20(4), 581–589. https://doi.org/10.1007/s10071-017-1082-x

Chance, M. R. A., & Mead, A. P. (1953). Social behaviour and primate evolution (p. 49). Clarendon Press/Oxford University Press.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-Scale Deceit: Deception as a Marker of Two-, Three-, and Four-Year-Olds' Early Theories of Mind. Child Development, 60(6), 1263. https://doi.org/10.2307/1130919

Chapman, B. B., Morrell, L. J., Tosh, C. R., & Krause, J. (2010). Behavioural consequences of sensory plasticity in guppies. Proceedings of the Royal Society B: Biological Sciences, 277(1686), 1395–1401. https://doi.org/10.1098/rspb.2009.2055

Cheney, D. L., & Seyfarth, R. M. (1990). How monkeys see the world: Inside the mind of another species (pp. x, 377). University of Chicago Press.

Chomsky, N. (1966). Cartesian linguistics: A chapter in the history of rationalist thought (1. ed). Cambridge Univ. Press.

Coon, D., Mitterer, J. O., & Martini, T. (2018). Psychology: Modules for active learning (14e [edition]). Cengage Learning.

Corballis, M. C. (2003). From hand to mouth: The origins of language (Second printing, and first paperback printing). Princeton University Press.

Cross, F. R., & Jackson, R. R. (2009). Cross-modality priming of visual and olfactory selective attention by a spider that feeds indirectly on vertebrate blood. Journal of Experimental Biology, 212(12), 1869–1875. https://doi.org/10.1242/jeb.028126

Cuskley, C., & Kirby, S. (2013). Synesthesia, Cross-Modality, and Language Evolution. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199603329.013.0043

Dakin, R., McCrossan, O., Hare, J. F., Montgomerie, R., & Amador Kane, S. (2016). Biomechanics of the Peacock's Display: How Feather Structure and Resonance Influence Multimodal Signaling. PLOS ONE, 11(4), e0152759. https://doi.org/10.1371/journal.pone.0152759

Dally, J. M., Emery, N. J., & Clayton, N. S. (2004). Cache protection strategies by western scrub–jays (Aphelocoma californica): Hiding food in the shade. Proceedings of the Royal Society of London. Series B: Biological Sciences, 271(suppl_6). https://doi.org/10.1098/rsbl.2004.0190

Dambly-Chaudière, C., Sapède, D., Soubiran, F., Decorde, K., Gompel, N., & Ghysen, A. (2003). The lateral line of zebrafish: A model system for the analysis of morphogenesis and neural development in vertebrates. Biology of the Cell, 95(9), 579–587. https://doi.org/10.1016/j.biolcel.2003.10.005

Darwin, C. (1872). The expression of the emotions in man and animals. John Murray. https://doi.org/10.1037/10001-000

Davidson, D. (1975). Thought and Talk. In S. D. Guttenplan & S. Guttenplan (A c. Di), Mind and Language (pp. 1975–1977). Clarendon Press.

Davidson, D. (1980). Essays on actions and events (1st ed). Clarendon Press ; Oxford University Press.

Davidson, D. (1982). Rational Animals. Dialectica, 36(4), 317–327.

Davidson, D. (1997). Seeing through Language. Royal Institute of Philosophy Supplement, 42, 15–27. https://doi.org/10.1017/S135824610001016X

Davidson, D. (2001). Subjective, Intersubjective, Objective: Philosophical Essays Volume 3. Clarendon Press.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation?

De Villiers, J. G. (2021). The Role(s) of Language in Theory of Mind. In M. Gilead & K. N. Ochsner (A c. Di), The Neural Basis of Mentalizing (pp. 423–448). Springer International Publishing. https://doi.org/10.1007/978-3-030-51890-5_21

De Vries, M. H., Barth, A. C. R., Maiworm, S., Knecht, S., Zwitserlood, P., & Flöel, A. (2010). Electrical Stimulation of Broca's Area Enhances Implicit Learning of an Artificial Grammar. Journal of Cognitive Neuroscience, 22(11), 2427–2436. https://doi.org/10.1162/jocn.2009.21385

De Vries, M. H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. Cognition, 107(2), 763–774. https://doi.org/10.1016/j.cognition.2007.09.002

Deng, Y., Wu, Q., & Weng, X. (2013). Unimodal and multimodal regions for logographic language processing in left ventral occipitotemporal cortex. Frontiers in Human Neuroscience, 7. https://doi.org/10.3389/fnhum.2013.00619

Dennett, D. C. (1978). Beliefs about beliefs. Behavioural and Brain Sciences, 1(4), 568–570. https://doi.org/10.1017/S0140525X00076664

Descartes, R., & Maclean, I. (2008). A discourse on the method of correctly conducting one's reason and seeking truth in the sciences. Oxford University Press.

Dretske, F. I. (2006). Minimal Rationality. In S. Hurley & M. Nudds (A c. Di), Rational Animals? Oxford University Press.

Dummett, M. (1993). The Seas of Language. Oxford University Press.

Dummett, M. (2010). The nature and future of philosophy. Columbia University Press.

Dunbar, R. I. M. (1996). Grooming, gossip, and the evolution of language (1st Harvard University Press paperback ed). Harvard University Press.

Dunlop, R. A., Cato, D. H., & Noad, M. J. (2010). Your attention please: Increasing ambient noise levels elicits a change in communication behaviour in humpback whales (Megaptera novaeangliae). Proceedings of the Royal Society B: Biological Sciences, 277(1693), 2521–2529. https://doi.org/10.1098/rspb.2009.2319

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young Children's Understanding of Other People's Feelings and Beliefs: Individual Differences and Their Antecedents. Child Development, 62(6), 1352. https://doi.org/10.2307/1130811

Duss, S. B., Oggier, S., Reber, T. P., & Henke, K. (2011). Formation of semantic associations between subliminally presented face-word pairs. Consciousness and Cognition, 20(3), 928–935. https://doi.org/10.1016/j.concog.2011.03.018

Ebert, S. (2020). Theory of mind, language, and reading: Developmental relations from early childhood to early adolescence. Journal of Experimental Child Psychology, 191, 104739. https://doi.org/10.1016/j.jecp.2019.104739

Emanuel Robinson, A., Hertzog, C., & Dunlosky, J. (2006). Aging, Encoding Fluency, and Metacognitive Monitoring. Aging, Neuropsychology, and Cognition, 13(3–4), 458–478. https://doi.org/10.1080/13825580600572983

Emery, N. J., & Clayton, N. S. (2001). Effects of experience and social context on prospective caching strategies by scrub jays. Nature, 414(6862), 443–446. https://doi.org/10.1038/35106560

Ferrigno, S. (2022). Sequencing, Artificial Grammar, and Recursion in Primates. In B. L. Schwartz & M. J. Beran (A c. Di), Primate Cognitive Studies (1ª ed., pp. 260–290). Cambridge University Press. https://doi.org/10.1017/9781108955836.011

Fodor, J. A. (1992). A theory of the child's theory of mind. Cognition, 44(3), 283–296. https://doi.org/10.1016/0010-0277(92)90004-2

Fodor, J. A. (2010). The language of thought (Digit. repr). Harvard Univ. Pr.

Fouts, R., Mills, S. T., & Goodall, J. (1997). Next of kin: My conversations with chimpanzees. Harper.

Frank, S. L., Bod, R., & Christiansen, M. H. (2012). How hierarchical is language use? Proceedings of the Royal Society B: Biological Sciences, 279(1747), 4522–4531. https://doi.org/10.1098/rspb.2012.1741

Frisch, K. von. (1967). The Dance Language and Orientation of Bees: Harvard University Press. https://doi.org/10.4159/harvard.9780674418776

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & Van Schaik, C. P. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations. Biological Reviews, 94(5), 1809–1829. https://doi.org/10.1111/brv.12535

Fröhlich, M., & van Schaik, C. P. (2018). The function of primate multimodal communication. Animal Cognition, 21(5), 619–629. https://doi.org/10.1007/s10071-018-1197-8

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. Cognitive Development, 10(4), 483–527. https://doi.org/10.1016/0885-2014(95)90024-1

Fuller, R. A., Warren, P. H., & Gaston, K. J. (2007). Daytime noise predicts nocturnal singing in urban robins. Biology Letters, 3(4), 368–370. https://doi.org/10.1098/rsbl.2007.0134

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain: A Journal of Neurology, 119 (Pt 2), 593–609. https://doi.org/10.1093/brain/119.2.593

Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. American Journal of Primatology, 2(3), 237–248. https://doi.org/10.1002/ajp.1350020302

Gaudet, I., Hüsser, A., Vannasing, P., & Gallagher, A. (2020). Functional Brain Connectivity of Language Functions in Children Revealed by EEG and MEG: A Systematic Review. Frontiers in Human Neuroscience, 14, 62. https://doi.org/10.3389/fnhum.2020.00062

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. Nature, 440(7088), 1204–1207. https://doi.org/10.1038/nature04675

Gernsbacher, M. A., & Yergeau, M. (2019). Empirical failures of the claim that autistic people lack a theory of mind. Archives of Scientific Psychology, 7(1), 102–118. https://doi.org/10.1037/arc0000067

Ghazanfar, A. A. (2013). Multisensory vocal communication in primates and the evolution of rhythmic speech. Behavioural Ecology and Sociobiology, 67(9), 1441–1448. https://doi.org/10.1007/s00265-013-1491-z

Ghazanfar, A. A., & Logothetis, N. K. (2003). Facial expressions linked to monkey calls. Nature, 423(6943), 937–938. https://doi.org/10.1038/423937a

Gleeson, R. A. (1991). Intrinsic Factors Mediating Pheromone Communication in the Blue Crab, Callinectes sapidus. In R. T. Bauer & J. W. Martin (A c. Di), Crustacean Sexual Biology (pp. 17–32). Columbia University Press. https://doi.org/10.7312/baue90796-003

Glock, H. J., Andrews, K., & Beck, J. (2018). Animal Rationality and Belief. In Glock, Hans Johann (2018). Animal rationality and belief. In: Andrews, Kirstin; Beck, Jacob. The Routledge Handbook of Philosophy of Animal Minds. London: Routledge, 89-99 (pp. 89–99).

Glock, H.-J. (2000). Animals, Thoughts and Concepts. Synthese, 123(1), 35–64. https://doi.org/10.1023/a:1005295521736

Glock, H.-J. (2010). Can Animals Judge? Dialectica, 64(1), 11–33. https://doi.org/10.1111/j.1746-8361.2010.01227.x

Godfrey-smith, P. (2016). Other minds: The Octopus, the Sea, and the Deep Origins of Consciousness.

Goodale, M., & Milner, D. (2013). Sight Unseen. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199596966.001.0001

Gopnik, A., & Meltzoff, A. N. (1997). Words, thoughts, and theories (pp. xvi, 268). The MIT Press.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (A c. Di), Mapping the Mind (1ª ed., pp. 257–293). Cambridge University Press. https://doi.org/10.1017/CBO9780511752902.011

Gordon, S. D., & Uetz, G. W. (2011). Multimodal communication of wolf spiders on different substrates: Evidence for behavioural plasticity. Animal Behaviour, 81(2), 367–375. https://doi.org/10.1016/j.anbehav.2010.11.003

Grafe, T. U., Preininger, D., Sztatecsny, M., Kasah, R., Dehling, J. M., Proksch, S., & Hödl, W. (2012). Multimodal Communication in a Noisy Environment: A Case Study of the Bornean Rock Frog Staurois parvus. PLoS ONE, 7(5), e37965. https://doi.org/10.1371/journal.pone.0037965

Grazzani, I., Ornaghi, V., Conte, E., Pepe, A., & Caprin, C. (2018). The Relation Between Emotion Understanding and Theory of Mind in Children Aged 3 to 8: The Key Role of Language. Frontiers in Psychology, 9, 724. https://doi.org/10.3389/fpsyg.2018.00724

Grice, H. P. (1957). Meaning. The Philosophical Review, 66(3), 377. https://doi.org/10.2307/2182440

Grice, H. P. (1975). Logic and Conversation. In D. Davidson & G. Harman (A c. Di), The Logic of Grammar (pp. 64–75).

Griffin, D. R. (1992). Animal Minds. University of Chicago Press.

Halina, M. (2015). There Is No Special Problem of Mindreading in Nonhuman Animals. Philosophy of Science, 82(3), 473–490. https://doi.org/10.1086/681627

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. Animal Behaviour, 59(4), 771–785. https://doi.org/10.1006/anbe.1999.1377

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? Animal Behaviour, 61(1), 139–151. https://doi.org/10.1006/anbe.2000.1518

Harman, G. (1978). Studying the chimpanzee's theory of mind. Behavioural and Brain Sciences, 1(4), 576–577. https://doi.org/10.1017/S0140525X00076743

Hauser, M. D. (1998). A nonhuman primate's expectations about object motion and destination: The importance of self-propelled movement and animacy. Developmental Science, 1(1), 31–37. https://doi.org/10.1111/1467-7687.00009

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? Science, 298(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hazlett, B. A., & Mclay, C. (2005). Responses to predation risk: Alternative strategies in the crab Heterozius rotundifrons. Animal Behaviour, 69(4), 967–972. https://doi.org/10.1016/j.anbehav.2004.06.028

Hebets, E. A. (2005). Attention-altering signal interactions in the multimodal courtship display of the wolf spider Schizocosa uetzi. Behavioural Ecology, 16(1), 75–82. https://doi.org/10.1093/beheco/arh133

Hebets, E. A. (2011). Current Status and Future Directions of Research in Complex Signaling. Current Zoology, 57(2), i–v. https://doi.org/10.1093/czoolo/57.2.i

Hebets, E. A., & Vink, C. J. (2007). Experience leads to preference: Experienced females prefer brush-legged males in a population of syntopic wolf spiders. Behavioural Ecology, 18(6), 1010–1020. https://doi.org/10.1093/beheco/arm070

Hempel, C. G. (1958). The theoretician's dilemma: A study in the logic of theory construction. Minnesota Studies in the Philosophy of Science, 2, 173–226.

Henke, K., Reber, T. P., & Duss, S. B. (2013). Integrating Events Across Levels of Consciousness. Frontiers in Behavioural Neuroscience, 7. https://doi.org/10.3389/fnbeh.2013.00068

Heyes, C. (2015). Animal mindreading: What’s the problem? Psychonomic Bulletin & Review, 22(2), 313–327. https://doi.org/10.3758/s13423-014-0704-4

Heyes, C. M. (1998). Theory of mind in nonhuman primates. Behavioural and Brain Sciences, 21(1), 101–114. https://doi.org/10.1017/S0140525X98000703

Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal communication. Behavioural Ecology and Sociobiology, 67(9), 1381–1388. https://doi.org/10.1007/s00265-013-1590-x

Hirata, S., & Matsuzawa, T. (2001). Tactics to obtain a hidden food item in chimpanzee pairs (Pan troglodytes). Animal Cognition, 4(3–4), 285–295. https://doi.org/10.1007/s100710100096

Hocking, J., & Price, C. J. (2008). The Role of the Posterior Superior Temporal Sulcus in Audiovisual Processing. Cerebral Cortex, 18(10), 2439–2449. https://doi.org/10.1093/cercor/bhn007

Hodl, Walter & Amézquita, Adolfo. (2001). Visual Signaling in Anuran Amphibians. In Anuran communication (pp. 121–141). Ryan, Micheal J. https://api.semanticscholar.org/CorpusID:202546511

Hoerold, D., Dockree, P. M., O’Keeffe, F. M., Bates, H., Pertl, M., & Robertson, I. H. (2008). Neuropsychology of self-awareness in young adults. Experimental Brain Research, 186(3), 509–515. https://doi.org/10.1007/s00221-008-1341-9

Hollenstein, N., Renggli, C., Glaus, B., Barrett, M., Troendle, M., Langer, N., & Zhang, C. (2021). Decoding EEG Brain Activity for Multi-Modal Natural Language Processing. Frontiers in Human Neuroscience, 15, 659410. https://doi.org/10.3389/fnhum.2021.659410

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), 337–343. https://doi.org/10.1037/0735-7036.115.4.337

Hostetter, A. B., Russell, J. L., Freeman, H., & Hopkins, W. D. (2006). Now you see me, now you don’t: Evidence that chimpanzees understand the role of the eyes in attention. Animal Cognition, 10(1), 55–62. https://doi.org/10.1007/s10071-006-0031-x

Hughes, M. (1996). The function of concurrent signals: Visual and chemical communication in snapping shrimp. Animal Behaviour, 52(2), 247–257. https://doi.org/10.1006/anbe.1996.0170

Hume, D. (2009). A treatise of human nature (D. F. Norton, A c. Di; Reprint with corrections). Oxford University Press.

Humphrey, N. K. (1976). The social function of intellect. In Growing points in ethology. Cambridge U Press.

Hurley, S., & Nudds, M. (2006). Rational animals? (pp. viii, 561). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528272.001.0001

Iwasa, Y., & Pomiankowski, A. (1994). The evolution of mate preferences for multiple sexual ornaments. Evolution, 48(3), 853–867. https://doi.org/10.1111/j.1558-5646.1994.tb01367.x

Jackson, R. R. (1977). Courtship versatility in the jumping spider, Phidippus johnsoni (Araneae: Salticidae). Animal Behaviour, 25, 953–957. https://doi.org/10.1016/0003-3472(77)90046-X

Jackson, R. R. (1992). Conditional strategies and interpopulation variation in the behaviour of jumping spiders. New Zealand Journal of Zoology, 19(3–4), 99–111. https://doi.org/10.1080/03014223.1992.10422314

Jaroš, F., & Pudil, M. (2020). Cognitive Systems of Human and Non-human Animals: At the Crossroads of Phenomenology, Ethology and Biosemiotics. Biosemiotics, 13(2), 155–177. https://doi.org/10.1007/s12304-020-09387-8

Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). Introducing Multimodality (0 ed.). Routledge. https://doi.org/10.4324/9781315638027

Johnson, C. N., & Wellman, H. M. (1982). Children's Developing Conceptions of the Mind and Brain. Child Development, 53(1), 222. https://doi.org/10.2307/1129656

Johnstone, R. A. (1995a). Honest advertisement of multiple qualities using multiple signals. Journal of Theoretical Biology, 177(1), 87–94. https://doi.org/10.1016/S0022-5193(05)80006-2

Johnstone, R. A. (1995b). Sexual selection, honest advertisement and the handicap principle: Reviewing the evidence. Biological Reviews, 70(1), 1–65. https://doi.org/10.1111/j.1469-185X.1995.tb01439.x

Johnstone, R. A., Reynolds, J. D., & Deutsch, J. C. (1996). Mutual mate choice and sex differences in choosiness. Evolution, 50(4), 1382–1391. https://doi.org/10.1111/j.1558-5646.1996.tb03912.x

Jolly, A. (1966). Lemur social behaviour and primate intelligence. Science, 153(3735), 501–506. https://doi.org/10.1126/science.153.3735.501

Jones, M. W., Branigan, H. P., Parra, M. A., & Logie, R. H. (2013). Cross-modal binding in developmental dyslexia. Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(6), 1807–1822. https://doi.org/10.1037/a0033334

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. Cognition, 109(2), 224–234. https://doi.org/10.1016/j.cognition.2008.08.010

Kamio, M., Reidenbach, M. A., & Derby, C. D. (2008). To paddle or not: Context dependent courtship display by male blue crabs, Callinectes sapidus. Journal of Experimental Biology, 211(8), 1243–1248. https://doi.org/10.1242/jeb.014977

Kasuya-Ueba, Y., Zhao, S., & Toichi, M. (2020). The Effect of Music Intervention on Attention in Children: Experimental Evidence. Frontiers in Neuroscience, 14, 757. https://doi.org/10.3389/fnins.2020.00757

Kellogg, R. T. (2016). Fundamentals of cognitive psychology (Third edition). SAGE.

Kirk, S. A. (2015). Educating exceptional children (14th edition). Cengage Learning.

Köhler, W. (1948). The mentality of apes, 1917. In Readings in the history of psychology (pp. 497–505). Appleton-Century-Crofts. https://doi.org/10.1037/11304-054

Krachun, C., Carpenter, C. M., Call, J., & Tomasello, M. (2010). A New Change-of-Contents False Belief Test: Children and Chimpanzees Compared. International Journal of Comparative Psychology, 23(2). https://doi.org/10.46867/IJCP.2010.23.02.03

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. Developmental Science, 12(4), 521–535. https://doi.org/10.1111/j.1467-7687.2008.00793.x

Ku, B., & Sung, M.-C. (2022). The Effects of Interventions on Motor Skills in Individuals with Down Syndrome: A Meta-Analysis. Journal of Developmental and Physical Disabilities, 34(5), 775–793. https://doi.org/10.1007/s10882-021-09827-4

Kunihira, Y., Senju, A., Dairoku, H., Wakabayashi, A., & Hasegawa, T. (2006). 'Autistic' Traits in Non-Autistic Japanese Populations: Relationships with Personality Traits and Cognitive Ability. Journal of Autism and Developmental Disorders, 36(4), 553–566. https://doi.org/10.1007/s10803-006-0094-1

Kuroshima, H., Fujita, K., Adachi, I., Iwata, K., & Fuyuki, A. (2003). A Capuchin monkey (Cebus apella) recognizes when people do and do not know the location of food. Animal Cognition, 6(4), 283–291. https://doi.org/10.1007/s10071-003-0184-9

Kuroshima, H., Fujita, K., Fuyuki, A., & Masuda, T. (2002). Understanding of the relationship between seeing and knowing by tufted capuchin monkeys (Cebus apella). Animal Cognition, 5(1), 41–48. https://doi.org/10.1007/s10071-001-0123-6

Larson, C., Thomas, H. R., Crutcher, J., Stevens, M. C., & Eigsti, I.-M. (2023). Language Networks in Autism Spectrum Disorder: A systematic review of connectivity-based fMRI studies. Review Journal of Autism and Developmental Disorders. https://doi.org/10.1007/s40489-023-00382-6

Lea, S. E. G. (1984). In What Sense do Pigeons Learn Concepts? In Animal Cognition. Psychology Press.

Leavens, D. A., & Hopkins, W. D. (1998). Intentional communication by chimpanzees: A cross-sectional study of the use of referential gestures. Developmental Psychology, 34(5), 813–822. https://doi.org/10.1037/0012-1649.34.5.813

Leavens, D. A., Hopkins, W. D., & Thomas, R. K. (2004a). Referential Communication by Chimpanzees (Pan troglodytes). Journal of Comparative Psychology, 118(1), 48–57. https://doi.org/10.1037/0735-7036.118.1.48

Leavens, D. A., Hostetter, A. B., Wesley, M. J., & Hopkins, W. D. (2004b). Tactical use of unimodal and bimodal communication by chimpanzees, Pan troglodytes. Animal Behaviour, 67(3), 467–476. https://doi.org/10.1016/j.anbehav.2003.04.007

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal communication by captive chimpanzees (Pan troglodytes). Animal Cognition, 13(1), 33–40. https://doi.org/10.1007/s10071-009-0242-z

Lee, D., Barraclough, D. J., & Seo, H. (1993). Neural basis of social interactions in primates. In P. Haggard, Y. Rossetti, & M. Kawato (A c. Di), Sensorimotor Foundations of Higher Cognition (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199231447.003.0012

Lee, L., Harkness, K. L., Sabbagh, M. A., & Jacobson, J. A. (2005). Mental state decoding abilities in clinical depression. Journal of Affective Disorders, 86(2–3), 247–258. https://doi.org/10.1016/j.jad.2005.02.007

Leonard, A. S., Dornhaus, A., & Papaj, D. R. (2011). Forget-me-not: Complex floral displays, inter-signal interactions, and pollinator cognition. Current Zoology, 57(2), 215–224. https://doi.org/10.1093/czoolo/57.2.215

Lev-Ari, T., Beeri, H., & Gutfreund, Y. (2022). The Ecological View of Selective Attention. Frontiers in Integrative Neuroscience, 16, 856207. https://doi.org/10.3389/fnint.2022.856207

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130302. https://doi.org/10.1098/rstb.2013.0302

Liaukovich, K., Ukraintseva, Y., & Martynova, O. (2022). Implicit auditory perception of local and global irregularities in passive listening condition. Neuropsychologia, 165, 108129. https://doi.org/10.1016/j.neuropsychologia.2021.108129

Lin, N., Yang, X., Li, J., Wang, S., Hua, H., Ma, Y., & Li, X. (2018). Neural correlates of three cognitive processes involved in theory of mind and discourse comprehension. Cognitive, Affective, & Behavioural Neuroscience, 18(2), 273–283. https://doi.org/10.3758/s13415-018-0568-6

Lorenz, K. (1935). Der Kumpan in der Umwelt des Vogels. Journal für Ornithologie, 83(2), 137–213. https://doi.org/10.1007/BF01905355

Loukusa, S., Mäkinen, L., Kuusikko-Gauffin, S., Ebeling, H., & Moilanen, I. (2014). Theory of mind and emotion recognition skills in children with specific language impairment, autism spectrum disorder and typical development: Group differences and connection to knowledge of grammatical morphology, word-finding abilities and verbal worki: Social perception in SLI, ASD and TD. International Journal of Language & Communication Disorders, 49(4), 498–507. https://doi.org/10.1111/1460-6984.12091

Lukito, S., Jones, C. R. G., Pickles, A., Baird, G., Happé, F., Charman, T., & Simonoff, E. (2017). Specificity of executive function and theory of mind performance in relation to attention-deficit/hyperactivity symptoms in autism spectrum disorders. Molecular Autism, 8(1), 60. https://doi.org/10.1186/s13229-017-0177-1

Luo, Y., & Baillargeon, R. (2005). Can a Self-Propelled Box Have a Goal?: Psychological Reasoning in 5-Month-Old Infants. Psychological Science, 16(8), 601–608. https://doi.org/10.1111/j.1467-9280.2005.01582.x

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? Cognition, 105(3), 489–512. https://doi.org/10.1016/j.cognition.2006.10.007

Lurz, R. (2009). If chimpanzees are mindreaders, could behavioural science tell? Toward a solution of the logical problem. Philosophical Psychology, 22(3), 305–328. https://doi.org/10.1080/09515080902970673

Lurz, R., & Krachun, C. (2019). Experience-Projection Methods in Theory-of-Mind Research: Their Limits and Strengths. Current Directions in Psychological Science, 28(5), 456–462. https://doi.org/10.1177/0963721419850156

Lurz, R. W. (A c. Di). (2009). The philosophy of animal minds. Cambridge University Press.

Lurz, R. W. (2011). Mindreading animals: The debate over what animals know about other minds. MIT Press.

Lurz, R. W., & Krachun, C. (2011b). How Could We Know Whether Nonhuman Primates Understand Others' Internal Goals and Intentions? Solving Povinelli's Problem. Review of Philosophy and Psychology, 2(3), 449–481. https://doi.org/10.1007/s13164-011-0068-x

Lyyra, P. (2005). Review of Jose Luis Bermudez: Thinking Without Words. PSYCHE: An Interdisciplinary Journal of Research On Consciousness, 11.

Ma, H. L., Dawson, M. R. W., Prinsen, R. S., & Hayward, D. A. (2023). Embodying cognitive ethology. Theory & Psychology, 33(1), 42–58. https://doi.org/10.1177/09593543221126165

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant Intermodal Speech Perception Is a Left-Hemisphere Function. Science, 219(4590), 1347–1349. https://doi.org/10.1126/science.6828865

Marino, L. (2002). Convergence of Complex Cognitive Abilities in Cetaceans and Primates. Brain, Behaviour and Evolution, 59(1–2), 21–32. https://doi.org/10.1159/000063731

Marler, P. (1967). Animal Communication Signals: We are beginning to understand how the structure of animal signals relates to the function they serve. Science, 157(3790), 769–774. https://doi.org/10.1126/science.157.3790.769

Marler, P. R. (1965). Communication in Monkeys and Apes.

Mash, E. J., & Wolfe, D. A. (2019). Abnormal child psychology (Seventh edition). Cengage.

Maynard Smith, J., & Harper, D. (2003). Animal signals (1st ed). Oxford University Press.

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. Nature, 264(5588), 746–748. https://doi.org/10.1038/264746a0

Meltzoff, A. N. (1999). Origins of theory of mind, cognition and communication. Journal of Communication Disorders, 32(4), 251–269. https://doi.org/10.1016/S0021-9924(99)00009-X

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. Nature, 282(5737), 403–404. https://doi.org/10.1038/282403a0

Meltzoff, A. N., & Brooks, R. (2008). Self-experience as a mechanism for learning about others: A training study in social cognition. Developmental Psychology, 44(5), 1257–1265. https://doi.org/10.1037/a0012888

Meltzoff, A. N., & Gopnik, A. (1993). The role of imitation in understanding persons and developing a theory of mind. In Understanding other minds (pp. 335–366). Understanding other minds.

Menzel, E. W. (1988). A group of young chimpanzees in a 1-acre field: Leadership and communication (p. 159). Clarendon Press/Oxford University Press.

Metcalfe, J., & Shimamura, A. P. (A c. Di). (1994). Metacognition: Knowing about Knowing. The MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001

Mill, J. S. (1843). A system of logic: Ratiocinative and inductive: being a connected view of the principles of evidence and the methods of scientific investigation (Repr). Univ. Press of the Pacific.

Millikan, R. G. (1996). Pushmi-pullyu representations (p. 161). The MIT Press.

Millikan, R. G. (2006). Styles of Rationality. In S. Hurley & M. Nudds (A c. Di), Rational Animals? Oxford University Press.

Mithen, S. (1990). Thoughtful foragers: A study of prehistoric decision making (1. publ). Cambridge University Press.

Mithen, S. J. (1999). The prehistory of the mind: The cognitive origins of art, religion and science (1st paperback ed). Thames and Hudson.

Møller, A. P. (1988). False Alarm Calls as a Means of Resource Usurpation in the Great Tit Parus major. Ethology, 79(1), 25–30. https://doi.org/10.1111/j.1439-0310.1988.tb00697.x

Møller, A. P., & Pomiankowski, A. (1993). Fluctuating asymmetry and sexual selection. Genetica, 89(1), 267–279. https://doi.org/10.1007/BF02424520

Moller, A. P., & Pomiankowski, A. (1993). Why have birds got multiple sexual ornaments? Behavioural Ecology and Sociobiology, 32(3). https://doi.org/10.1007/BF00173774

Monclús, R., Arroyo, M., Valencia, A., & De Miguel, F. J. (2009). Red foxes (Vulpes vulpes) use rabbit (Oryctolagus cuniculus) scent marks as territorial marking sites. Journal of Ethology, 27(1), 153–156. https://doi.org/10.1007/s10164-008-0098-8

Monclús, R., Palomares, F., Tablado, Z., Martínez-Fontúrbel, A., & Palme, R. (2009). Testing the threat-sensitive predator avoidance hypothesis: Physiological responses and predator pressure in wild rabbits. Oecologia, 158(4), 615–623. https://doi.org/10.1007/s00442-008-1201-0

Morgan, C. L. (1894). An introduction to comparative psychology. Walter Scott Publishing Co. https://doi.org/10.1037/11344-000

Munakata, Y., & Yerys, B. E. (2001). All Together Now: When Dissociations Between Knowledge and Action Disappear. Psychological Science, 12(4), 335–337. https://doi.org/10.1111/1467-9280.00361

Munn, C. A. (1986). Birds that 'cry wolf'. Nature, 319(6049), 143–145. https://doi.org/10.1038/319143a0

Munoz, N. E., & Blumstein, D. T. (2012). Multisensory perception in uncertain environments. Behavioural Ecology, 23(3), 457–462. https://doi.org/10.1093/beheco/arr220

Murillo, E., Camacho, L., & Montero, I. (2021a). Multimodal Communication in Children with Autism Spectrum Disorder and Different Linguistic Development. Journal of Autism and Developmental Disorders, 51(5), 1528–1539. https://doi.org/10.1007/s10803-020-04637-7

Murillo, E., & Casla, M. (2021). Multimodal representational gestures in the transition to multi-word productions. Infancy, 26(1), 104–122. https://doi.org/10.1111/infa.12375

Murillo, E., Montero, I., & Casla, M. (2021b). On the Multimodal Path to Language: The Relationship Between Rhythmic Movements and Deictic Gestures at the End of the First Year. Frontiers in Psychology, 12, 616812. https://doi.org/10.3389/fpsyg.2021.616812

Newen, A., & Starzak, T. (2022). How to ascribe beliefs to animals. Mind & Language, 37(1), 3–21. https://doi.org/10.1111/mila.12302

Norbury, C. Frazier. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. British Journal of Developmental Psychology, 23(3), 383–399. https://doi.org/10.1348/026151005X26732

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? Science, 308(5719), 255–258. https://doi.org/10.1126/science.1107621

Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. Acta Psychologica, 124(1), 106–128. https://doi.org/10.1016/j.actpsy.2006.09.009

Opoku-Baah, C., Schoenhaut, A. M., Vassall, S. G., Tovar, D. A., Ramachandran, R., & Wallace, M. T. (2021). Visual Influences on Auditory Behavioural, Neural, and Perceptual Processes: A

Review. Journal of the Association for Research in Otolaryngology, 22(4), 365–386. https://doi.org/10.1007/s10162-021-00789-0

Otovic, P., & Partan, S. (2009). Multimodal Signaling in Animals. In Encyclopedia of Neuroscience (pp. 1095–1105). Elsevier. https://doi.org/10.1016/B978-008045046-9.01860-X

Owren, M. J., Rendall, D., & Ryan, M. J. (2010). Redefining animal signaling: Influence versus information in communication. Biology & Philosophy, 25(5), 755–780. https://doi.org/10.1007/s10539-010-9224-4

Özaltun, E. (2021). Dummett ve Davidson: Düşüncenin Dile Bağlılığı Üzerine. Beytulhikme An International Journal of Philosophy, 11(11:1), 47–78. https://doi.org/10.18491/beytulhikme.1626

Partan, S., & Marler, P. (1999). Communication Goes Multimodal. Science, 283(5406), 1272–1273. https://doi.org/10.1126/science.283.5406.1272

Partan, S. R. (2013). Ten unanswered questions in multimodal communication. Behavioural Ecology and Sociobiology, 67(9), 1523–1539. https://doi.org/10.1007/s00265-013-1565-y

Partan, S. R. (2017). Multimodal shifts in noise: Switching channels to communicate through rapid environmental change. Animal Behaviour, 124, 325–337. https://doi.org/10.1016/j.anbehav.2016.08.003

Partan, S. R., Fulmer, A. G., Gounard, M. A. M., & Redmond, J. E. (2010). Multimodal alarm behaviour in urban and rural gray squirrels studied by means of observation and a mechanical robot. Current Zoology, 56(3), 313–326. https://doi.org/10.1093/czoolo/56.3.313

Partan, S. R., & Marler, P. (2002). The Umwelt and its relevance to animal communication: Introduction to special issue. Journal of Comparative Psychology, 116(2), 116–119. https://doi.org/10.1037/0735-7036.116.2.116

Partan, S. R., & Marler, P. (2005). Issues in the Classification of Multimodal Communication Signals. The American Naturalist, 166(2), 231–245. https://doi.org/10.1086/431246

Paunov, A. M., Blank, I. A., Jouravlev, O., Mineroff, Z., Gallée, J., & Fedorenko, E. (2022). Differential Tracking of Linguistic vs. Mental State Content in Naturalistic Stimuli by Language and Theory of Mind (ToM) Brain Networks. Neurobiology of Language, 3(3), 413–440. https://doi.org/10.1162/nol_a_00071

Payne, R. S., & McVay, S. (1971). Songs of Humpback Whales: Humpbacks emit sounds in long, predictable patterns ranging over frequencies audible to humans. Science, 173(3997), 585–597. https://doi.org/10.1126/science.173.3997.585

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1480), 731–744. https://doi.org/10.1098/rstb.2006.2023

Penn, D. L., Sanna, L. J., & Roberts, D. L. (2008). Social Cognition in Schizophrenia: An Overview. Schizophrenia Bulletin, 34(3), 408–411. https://doi.org/10.1093/schbul/sbn014

Perner, J. (2000). Memory and Theory of Mind. In E. Tulving (A c. Di), The Oxford Handbook of Memory (pp. 297–312). Oxford University Press.

Perner, J. (2012). MiniMeta: In search of minimal criteria for metacognition. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (A c. Di), Foundations of Metacognition (pp. 94–116). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199646739.003.0007

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J., & Frith, C. D. (2008). Subliminal Instrumental Conditioning Demonstrated in the Human Brain. Neuron, 59(4), 561–567. https://doi.org/10.1016/j.neuron.2008.07.005

Peterson, C. C., & Siegal, M. (1995). Deafness, Conversation and Theory of Mind. Journal of Child Psychology and Psychiatry, 36(3), 459–474. https://doi.org/10.1111/j.1469-7610.1995.tb01303.x

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in Theory-of-Mind Development for Children With Deafness or Autism. Child Development, 76(2), 502–517. https://doi.org/10.1111/j.1467-8624.2005.00859.x

Plotnik, J. M., De Waal, F. B. M., & Reiss, D. (2006). Self-recognition in an Asian elephant. Proceedings of the National Academy of Sciences, 103(45), 17053–17057. https://doi.org/10.1073/pnas.0608062103

Povinelli, D. J. (1999). Social understanding in chimpanzees: New evidence from a longitudinal approach. In Developing theories of intention: Social understanding and self-control (pp. 195–225). Lawrence Erlbaum Associates Publishers.

Povinelli, D. J., Eddy, T. J., Hobson, R. P., & Tomasello, M. (1996). What Young Chimpanzees Know about Seeing. Monographs of the Society for Research in Child Development, 61(3), i. https://doi.org/10.2307/1166159

Povinelli, D. J., Nelson, K. E., & Boysen, S. T. (1990). Inferences about guessing and knowing by chimpanzees (Pan troglodytes). Journal of Comparative Psychology, 104(3), 203–210. https://doi.org/10.1037/0735-7036.104.3.203

Povinelli, D. J., Rulf, A. B., & Bierschwale, D. T. (1994). Absence of knowledge attribution and self-recognition in young chimpanzees (Pan troglodytes). Journal of Comparative Psychology, 108(1), 74–80. https://doi.org/10.1037/0735-7036.108.1.74

Povinelli, D. J., & Vonk, J. (2004). We Don’t Need a Microscope to Explore the Chimpanzee’s Mind. Mind and Language, 19(1), 1–28. https://doi.org/10.1111/j.1468-0017.2004.00244.x

Preininger, D., Boeckle, M., Freudmann, A., Starnberger, I., Sztatecsny, M., & Hödl, W. (2013). Multimodal signaling in the Small Torrent Frog (Micrixalus saxicola) in a complex acoustic environment. Behavioural Ecology and Sociobiology, 67(9), 1449–1456. https://doi.org/10.1007/s00265-013-1489-6

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? Behavioural and Brain Sciences, 1(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Proust, J. (2007). Metacognition and Metarepresentation: Is a Self-Directed Theory of Mind a Precondition for Metacognition? Synthese, 159(2), 271–295. https://doi.org/10.1007/s11229-007-9208-3

Putnam, H. (1991). Representation and Reality. The MIT Press. https://doi.org/10.7551/mitpress/5891.001.0001

Putnam, Hilary. (1976). Mind, language and reality. Philosophical papers, vol, 2. Cambridge: University Press, 1975. - Claus Mueller, The politics of communication: A study in the political sociology of language, socialisation and legitimation. London: Oxford University Press, 1973. Pp. X + 226. (Also in paperback.)—Americo Paredes & Richard Bauman (eds), Toward new perspectives in folklore. Austin, Texas and London: The University of Texas Press. Pp. 191 + xv. Language in Society, 5(2), 254–255. https://doi.org/10.1017/S0047404500007132

Quine, W. V. O. (1992). Word and object. Massachusetts institut of technology press.

Rabin, L. A., Coss, R. G., & Owings, D. H. (2006). The effects of wind turbines on antipredator behaviour in California ground squirrels (Spermophilus beecheyi). Biological Conservation, 131(3), 410–420. https://doi.org/10.1016/j.biocon.2006.02.016

Raio, C. M., Carmel, D., Carrasco, M., & Phelps, E. A. (2012). Nonconscious fear is quickly acquired but swiftly forgotten. Current Biology, 22(12), R477–R479. https://doi.org/10.1016/j.cub.2012.04.023

Regolin, L., & Loconsole, M. (2023). Behavioural Methods to Study Cognitive Capacities of Animals. Animals, 13(22), 3445. https://doi.org/10.3390/ani13223445

Reiss, D., & Marino, L. (2001). Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. Proceedings of the National Academy of Sciences, 98(10), 5937–5942. https://doi.org/10.1073/pnas.101086398

Rendall, D., Owren, M. J., & Ryan, M. J. (2009). What do animal signals mean? Animal Behaviour, 78(2), 233–240. https://doi.org/10.1016/j.anbehav.2009.06.007

Rescorla, M. (2009). Cognitive Maps and the Language of Thought. British Journal for the Philosophy of Science, 60(2), 377–407. https://doi.org/10.1093/bjps/axp012

Richert, R., & Harris, P. (2006). The Ghost in My Body: Children’s Developing Concept of the Soul. Journal of Cognition and Culture, 6(3–4), 409–427. https://doi.org/10.1163/156853706778554913

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. Cognitive Brain Research, 3(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0

Roche, M. (2013). Povinelli’s Problem and Introspection. Review of Philosophy and Psychology, 4(4), 559–576. https://doi.org/10.1007/s13164-013-0152-5

Rogers, J. (2023). Flexible and generalizable representations of touch. Nature Reviews Neuroscience, 24(3), 132–132. https://doi.org/10.1038/s41583-023-00679-9

Ronacher, B. (2019). Innate releasing mechanisms and fixed action patterns: Basic ethological concepts as drivers for neuroethological studies on acoustic communication in Orthoptera. Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioural Physiology, 205(1), 33–50. https://doi.org/10.1007/s00359-018-01311-3

Ronald, A., Viding, E., Happé, F., & Plomin, R. (2006). Individual differences in theory of mind ability in middle childhood and links with verbal ability and autistic traits: A twin study. Social Neuroscience, 1(3–4), 412–425. https://doi.org/10.1080/17470910601068088

Ronald, K. L., Fernández-Juricic, E., & Lucas, J. R. (2012). Taking the sensory approach: How individual differences in sensory perception can influence mate choice. Animal Behaviour, 84(6), 1283–1294. https://doi.org/10.1016/j.anbehav.2012.09.015

Rosello, B., Berenguer, C., Baixauli, I., García, R., & Miranda, A. (2020). Theory of Mind Profiles in Children With Autism Spectrum Disorder: Adaptive/Social Skills and Pragmatic Competence. Frontiers in Psychology, 11, 567401. https://doi.org/10.3389/fpsyg.2020.567401

Rosenthal, G. G., & Ryan, M. J. (2000). Visual and acoustic communication in non-human animals: A comparison. Journal of Biosciences, 25(3), 285–290. https://doi.org/10.1007/BF02703937

Ross, K. M., Bard, K. A., & Matsuzawa, T. (2014). Playful expressions of one-year-old chimpanzee infants in social and solitary play contexts. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.00741

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. Animal Behaviour, 58(5), 921–931. https://doi.org/10.1006/anbe.1999.1242

Rowe, C., & Guilford, T. (1999). Novelty effects in a multimodal warning signal. Animal Behaviour, 57(2), 341–346. https://doi.org/10.1006/anbe.1998.0974

Rowe, D. C. (1999). Introduction to the Special Section on Behavioural Genetics. International Journal of Behavioural Development, 23(2), 289–292. https://doi.org/10.1080/016502599383829

Rump, G. C., & Smith, W. J. (1979). The Behaviour of Communicating. An Ethological Approach. Leonardo, 12(1), 74. https://doi.org/10.2307/1574103

Rundus, A. S., Owings, D. H., Joshi, S. S., Chinn, E., & Giannini, N. (2007). Ground squirrels use an infrared signal to deter rattlesnake predation. Proceedings of the National Academy of Sciences, 104(36), 14372–14376. https://doi.org/10.1073/pnas.0702599104

Russell, J. L., Braccini, S., Buehler, N., Kachin, M. J., Schapiro, S. J., & Hopkins, W. D. (2005). Chimpanzee (Pan troglodytes) intentional communication is not contingent upon food. Animal Cognition, 8(4), 263–272. https://doi.org/10.1007/s10071-005-0253-3

Ruxton, G. D., & Schaefer, H. M. (2011). Resolving current disagreements and ambiguities in the terminology of animal communication: Definitions in animal communication. Journal of Evolutionary Biology, 24(12), 2574–2585. https://doi.org/10.1111/j.1420-9101.2011.02386.x

Saito, D. N., Yoshimura, K., Kochiyama, T., Okada, T., Honda, M., & Sadato, N. (2005). Cross-modal Binding and Activated Attentional Networks during Audio-visual Speech Integration: A Functional MRI Study. Cerebral Cortex, 15(11), 1750–1760. https://doi.org/10.1093/cercor/bhi052

Savage-Rumbaugh, E. S., & Lewin, R. (1994). Kanzi: The ape at the brink of the human mind. John Wiley & Sons.

Savage-Rumbaugh, E. S., Rumbaugh, D. M., & Boysen, S. (1978). Sarah’s Problems of Comprehension. Behavioural and Brain Sciences, 1(4), 555–557. https://doi.org/10.1017/s0140525x0007655x

Schaefer, H. M., & Ruxton, G. D. (2011). Plant-Animal Communication. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199563609.001.0001

Scheffer, S. J., Uetz, G. W., & Stratton, G. E. (1996). Sexual selection, male morphology, and the efficacy of courtship signalling in two wolf spiders (Araneae: Lycosidae). Behavioural Ecology and Sociobiology, 38(1), 17–23. https://doi.org/10.1007/s002650050212

Schluter, D., & Price, T. (1993). Honesty, perception and population divergence in sexually selected traits. Proceedings of the Royal Society of London. Series B: Biological Sciences, 253(1336), 117–122. https://doi.org/10.1098/rspb.1993.0089

Schwitzgebel, E. (2015). Belief. In E. N. Zalta & U. Nodelman (A c. Di), The Stanford Encyclopedia of Philosophy (Summer 2015). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2015/entries/belief/

Scott, R. B., Samaha, J., Chrisley, R., & Dienes, Z. (2018). Prevailing theories of consciousness are challenged by novel cross-modal associations acquired between subliminal stimuli. Cognition, 175, 169–185. https://doi.org/10.1016/j.cognition.2018.02.008

Secondi, J., Rodgers, G., Bayle, F., Sourice, S., & Théry, M. (2015). Mate preference, species recognition and multimodal communication in heterogeneous environments. Evolutionary Ecology, 29(2), 217–227. https://doi.org/10.1007/s10682-014-9744-5

Seed, A., & Byrne, R. (2010). Animal Tool-Use. Current Biology, 20(23), R1032–R1039. https://doi.org/10.1016/j.cub.2010.09.042

Seitz, A. R., Kim, D., & Watanabe, T. (2009). Rewards Evoke Learning of Unconsciously Processed Visual Stimuli in Adult Humans. Neuron, 61(5), 700–707. https://doi.org/10.1016/j.neuron.2009.01.016

Seyfarth, R. M., & Cheney, D. L. (1990). The assessment by vervet monkeys of their own and another species' alarm calls. Animal Behaviour, 40(4), 754–764. https://doi.org/10.1016/S0003-3472(05)80704-3

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980a). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. Science (New York, N.Y.), 210(4471), 801–803. https://doi.org/10.1126/science.7433999

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980b). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. Animal Behaviour, 28(4), 1070–1094. https://doi.org/10.1016/S0003-3472(80)80097-2

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. Developmental Psychology, 47(5), 1239–1247. https://doi.org/10.1037/a0023899

Shanks, D. R. (2010). Learning: From Association to Cognition. Annual Review of Psychology, 61(1), 273–301. https://doi.org/10.1146/annurev.psych.093008.100519

Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E., & Weaver, W. (1998). The mathematical theory of communication. University of Illinois Press.

Shettleworth, S. J. (1998). Cognition, evolution, and behaviour (pp. xv, 688). Oxford University Press.

Sigelman, C. K., & Rider, E. A. (2018). Life-span human development (9e, Student Edition ed.). Cengage Learning.

Siniscalchi, M., d'Ingeo, S., Minunno, M., & Quaranta, A. (2018). Communication in Dogs. Animals, 8(8), 131. https://doi.org/10.3390/ani8080131

Smith, H. A. (1979). Nonverbal Communication in Teaching. Review of Educational Research, 49(4), 631–672. https://doi.org/10.3102/00346543049004631

Soper, H., & Murray, MO. (2012). Autism. In C. A. Noggle, R. S. Dean, & A. M. Horton (A c. Di), The encyclopedia of neuropsychological disorders (pp. 125–128). Springer Pub.

Spelke, E. S. (1990). Principles of Object Perception. Cognitive Science, 14(1), 29–56. https://doi.org/10.1207/s15516709cog1401_3

Stafstrom, J. A., & Hebets, E. A. (2013). Female mate choice for multimodal courtship and the importance of the signaling background for selection on male ornamentation. Current Zoology, 59(2), 200–209. https://doi.org/10.1093/czoolo/59.2.200

Stich, S. P. (1979). Do Animals Have Beliefs? Australasian Journal of Philosophy, 57(1), 15–28. https://doi.org/10.1080/00048407912341011

Stratron, G. E., & Uetz, G. W. (1981). Acoustic Communication and Reproductive Isolation in Two Species of Wolf Spiders. Science, 214(4520), 575–577. https://doi.org/10.1126/science.214.4520.575

Stratton, G. E., & Uetz, G. W. (1983). Communication via substratum-coupled stridulation and reproductive isolation in wolf spiders (Araneae: Lycosidae). Animal Behaviour, 31(1), 164–172. https://doi.org/10.1016/S0003-3472(83)80185-7

Strauß, J., & Stumpner, A. (2015). Selective forces on origin, adaptation and reduction of tympanal ears in insects. Journal of Comparative Physiology A, 201(1), 155–169. https://doi.org/10.1007/s00359-014-0962-7

Strawson, P. (1959). Individuals: An Essay in Descriptive Metaphysics (W. Wang, A c. Di). Routledge.

Stulp, G., Emery, N. J., Verhulst, S., & Clayton, N. S. (2009). Western scrub-jays conceal auditory information when competitors can hear but cannot see. Biology Letters, 5(5), 583–585. https://doi.org/10.1098/rsbl.2009.0330

Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs. The Journal of the Acoustical Society of America, 119(3), 1849–1866. https://doi.org/10.1121/1.2161827

Swaisgood, R. R., Rowe, M. P., & Owings, D. H. (1999). Assessment of rattlesnake dangerousness by California ground squirrels: Exploitation of cues from rattling sounds. Animal Behaviour, 57(6), 1301–1310. https://doi.org/10.1006/anbe.1999.1095

Tachibana, R., & Noguchi, Y. (2015). Unconscious analyses of visual scenes based on feature conjunctions. Journal of Experimental Psychology: Human Perception and Performance, 41(3), 639–648. https://doi.org/10.1037/xhp0000045

Taglialatela, J. P., Russell, J. L., Pope, S. M., Morton, T., Bogart, S., Reamer, L. A., Schapiro, S. J., & Hopkins, W. D. (2015). Multimodal communication in chimpanzees: Multimodal Communication in Chimpanzees. American Journal of Primatology, 77(11), 1143–1148. https://doi.org/10.1002/ajp.22449

Taylor, P. W., Roberts, J. A., & Uetz, G. W. (2005). Flexibility in the multi-modal courtship of a wolf spider, Schizocosa ocreata. Journal of Ethology, 23(1), 71–75. https://doi.org/10.1007/s10164-004-0129-z

Taylor, R. C., Klein, B. A., Stein, J., & Ryan, M. J. (2008). Faux frogs: Multimodal signalling and the value of robotics in animal behaviour. Animal Behaviour, 76(3), 1089–1097. https://doi.org/10.1016/j.anbehav.2008.01.031

Taylor, S. V., & Leung, C. B. (2020). Multimodal Literacy and Social Interaction: Young Children’s Literacy Learning. Early Childhood Education Journal, 48(1), 1–10. https://doi.org/10.1007/s10643-019-00974-0

Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. British Journal of

Developmental Psychology, 30(1), 172–187. https://doi.org/10.1111/j.2044-835X.2011.02067.x

Thompson, J. T., Bissell, A. N., & Martins, E. P. (2008). Inhibitory interactions between multimodal behavioural responses may influence the evolution of complex signals. Animal Behaviour, 76(1), 113–121. https://doi.org/10.1016/j.anbehav.2007.12.015

Tinbergen, N. (1953a). The evolution of mating behaviour patterns. Evolution, 7(4), 391–391. https://doi.org/10.1111/j.1558-5646.1953.tb00102.x

Tinbergen, N. (1953b). The Herring Gull's world: A study of the social behaviour of birds. Collins.

Tinbergen, N. (1959a). BEHAVIOUR, SYSTEMATICS, AND NATURAL SELECTION. Ibis, 101(3–4), 318–330. https://doi.org/10.1111/j.1474-919X.1959.tb02388.x

Tinbergen, N. (1959b). Comparative Studies of the Behaviour of Gulls (Laridae): A Progress Report. Behaviour, 15(1–2), 1–69. https://doi.org/10.1163/156853960X00098

Tinbergen, N. (1963). On aims and methods of Ethology. Zeitschrift Für Tierpsychologie, 20(4), 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Tomasello, M. (2008). Origins of human communication (pp. xiii, 393). MIT Press.

Tomasello, M. (2023). Social cognition and metacognition in great apes: A theory. Animal Cognition, 26(1), 25–35. https://doi.org/10.1007/s10071-022-01662-0

Tomasello, M., & Call, J. (1997). Primate cognition. Oxford University Press.

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(2), 137–154. https://doi.org/10.1007/BF02382050

Tomasello, M., George, B. L., Kruger, A. C., Jeffrey, M., Farrar, & Evans, A. (1985). The development of gestural communication in young chimpanzees. Journal of Human Evolution, 14(2), 175–186. https://doi.org/10.1016/S0047-2484(85)80005-1

Traniello, I. M., & Robinson, G. E. (2021). Neural and Molecular Mechanisms of Biological Embedding of Social Interactions. Annual Review of Neuroscience, 44(1), 109–128. https://doi.org/10.1146/annurev-neuro-092820-012959

Trenholme, R., Fodor, J., & Lepore, E. (1994). Holism: A Shopper's Guide. Noûs, 28(2), 241. https://doi.org/10.2307/2216051

Tschudin, A. J.-P. C. (2006). Belief attribution tasks with dolphins: What social minds can reveal about animal rationality. In S. Hurley & M. Nudds (A c. Di), Rational Animals? (pp. 413–436). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528272.003.0019

Tucci, S. L., Easterbrooks, S. R., & Lederberg, A. R. (2016). The Effects of Theory of Mind Training on the False Belief Understanding of Deaf and Hard-of-Hearing Students in Prekindergarten and Kindergarten. Journal of Deaf Studies and Deaf Education, 21(3), 310–325. https://doi.org/10.1093/deafed/enw031

Tye, M. (1997). The Problem of Simple Minds: Is There Anything It Is like to Be a Honey Bee? Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 88(3), 289–317.

Uetz, G. W., & Stratton, G. (1983). Communication in spiders. Endeavour, 7(1), 13–18. https://doi.org/10.1016/0160-9327(83)90042-X

Valle, C. L., Chenausky, K., & Tager-Flusberg, H. (2021). How do minimally verbal children and adolescents with autism spectrum disorder use communicative gestures to complement their spoken language abilities? Autism & Developmental Language Impairments, 6, 239694152110350. https://doi.org/10.1177/23969415211035065

van der Sluijs, I., Gray, S. M., Amorim, M. C. P., Barber, I., Candolin, U., Hendry, A. P., Krahe, R., Maan, M. E., Utne-Palm, A. C., Wagner, H.-J., & Wong, B. B. M. (2011). Communication in troubled waters: Responses of fish communication systems to changing environments. Evolutionary Ecology, 25(3), 623–640. https://doi.org/10.1007/s10682-010-9450-x

Verheggen, C. (A c. Di). (2017). Wittgenstein and Davidson on Language, Thought, and Action (1ª ed.). Cambridge University Press. https://doi.org/10.1017/9781316145364

Veroude, K., Norris, D. G., Shumskaya, E., Gullberg, M., & Indefrey, P. (2010). Functional connectivity between brain regions involved in learning words of a new language. Brain and Language, 113(1), 21–27. https://doi.org/10.1016/j.bandl.2009.12.005

Vigo, P. R., & Allen, C. (2009). How to Reason Without Words: Inference as Categorization. Cognitive Processing, 10, 77–88.

Vonk, D. P. & & Jennifer. (2006). We Don’t Need a Microscope to Explore the Chimpanzee’s Mind. In S. Hurley & M. Nudds (A c. Di), Rational Animals? Oxford University Press.

Vonk, J. (2016). Unobservability Hypothesis, The (Vonk and Povinelli, 2006). In V. Weekes-Shackelford, T. K. Shackelford, & V. A. Weekes-Shackelford (A c. Di), Encyclopedia of Evolutionary Psychological Science (pp. 1–3). Springer International Publishing. https://doi.org/10.1007/978-3-319-16999-6_3115-1

Vonk, J., & Povinelli, D. J. (2006). Similarity and Difference in the Conceptual Systems of Primates: The Unobservability Hypothesis. In Comparative cognition: Experimental explorations of animal intelligence (pp. 363–387). Oxford University Press.

Wade, M., Prime, H., Jenkins, J. M., Yeates, K. O., Williams, T., & Lee, K. (2018). On the relation between theory of mind and executive functioning: A developmental cognitive neuroscience perspective. Psychonomic Bulletin & Review, 25(6), 2119–2140. https://doi.org/10.3758/s13423-018-1459-0

Walton, G. E., & Bower, T. G. R. (1993). Newborns Form “Prototypes” in Less Than 1 Minute. Psychological Science, 4(3), 203–205. https://doi.org/10.1111/j.1467-9280.1993.tb00488.x

Wang, M. V., Lekhal, R., Aaro, L. E., Holte, A., & Schjolberg, S. (2014). The developmental relationship between language and motor performance from 3 to 5 years of age: A prospective longitudinal population study. BMC Psychology, 2(1), 34. https://doi.org/10.1186/s40359-014-0034-3

Wang, S., Allen, R. J., Fang, S.-Y., & Li, P. (2017). Cross-modal working memory binding and L1-L2 word learning. Memory & Cognition, 45(8), 1371–1383. https://doi.org/10.3758/s13421-017-0731-2

Wang, Y., Wang, Y., Chen, S., Zhu, C., & Wang, K. (2008). Theory of mind disability in major depression with or without psychotic symptoms: A componential view. Psychiatry Research, 161(2), 153–161. https://doi.org/10.1016/j.psychres.2007.07.018

Washburn, D., Wilson, G., Roes, M., Rnic, K., & Harkness, K. L. (2016). Theory of mind in social anxiety disorder, depression, and comorbid conditions. Journal of Anxiety Disorders, 37, 71–77. https://doi.org/10.1016/j.janxdis.2015.11.004

Welker, D. (1988). On the Necessity of Bodies. Erkenntnis, 28(3), 363–385. https://doi.org/10.1007/bf00184901

Wellman, H. M. (1990). The child's theory of mind. MIT Press.

Wellman, H. M. (2017). The Development of Theory of Mind: Historical Reflections. Child Development Perspectives, 11(3), 207–214. https://doi.org/10.1111/cdep.12236

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. Cognition, 30(3), 239–277. https://doi.org/10.1016/0010-0277(88)90021-2

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of Theory-of-Mind Understandings in Chinese Children. Psychological Science, 17(12), 1075–1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. Child Development, 75(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Whiten, A., & Byrne, R. W. (1988). Tactical Deception in Primates. Behavioural and Brain Sciences, 11(2), 233–244. https://doi.org/10.1017/s0140525x00049682

Whiten, A., & Byrne, R. W. (1997). Machiavellian intelligence II: Extensions and evaluations (pp. xii, 403). Cambridge University Press. https://doi.org/10.1017/CBO9780511525636

Wilgers, D. J., & Hebets, E. A. (2011). Complex courtship displays facilitate male reproductive success and plasticity in signaling across variable environments. Current Zoology, 57(2), 175–186. https://doi.org/10.1093/czoolo/57.2.175

Wimmer, H. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition, 13(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Wittgenstein, L. (1953). Philosophical investigations. Philosophische Untersuchungen (pp. x, 232). Macmillan.

Wolkenstein, L., Schönenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. Journal of Affective Disorders, 132(1–2), 104–111. https://doi.org/10.1016/j.jad.2011.02.010

Wong, B. B. M. (2014). Animal communication in a human-dominated world: A comment on Radford et al. Behavioural Ecology, 25(5), 1033–1033. https://doi.org/10.1093/beheco/aru139

Xia, R. J., & Haas, B. W. (2023). The effect of bilingualism and multicultural experience on social-cognitive processing: A meta-analytic review. Journal of Cultural Cognitive Science. https://doi.org/10.1007/s41809-023-00138-y

Zhao, C., Shang, S., Compton, A. M., Fu, G., & Sai, L. (2021). A Longitudinal Study of the Relations Between Theory of Mind, Executive Function, and Lying in Children. Frontiers in Psychology, 12, 766891. https://doi.org/10.3389/fpsyg.2021.766891

Zuberbühler, K., Cheney, D. L., & Seyfarth, R. M. (1999). Conceptual semantics in a nonhuman primate. Journal of Comparative Psychology, 113(1), 33–42. https://doi.org/10.1037/0735-7036.113.1.33